An evaluation of the use of data mining in counter-terrorism

Moeckli, Daniel ; Thurman, James

FP7-SECT-2007-217862

DETECTER
Detection Technologies, Terrorism, Ethics and Human Rights

Collaborative Project

An Evaluation of the Use of Data Mining in Counter-Terrorism
D08.2

Due date of deliverable: November 2010
Actual submission date: February 2011

Start date of project: 1.12.2008 Duration: 36 months

Work Package number and lead: WP06 Dr. Daniel Moeckli
Author(s): Dr. Daniel Moeckli, University of Zurich; James Thurman, University of Zurich

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	**x**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission	
CO	Confidential, only for members of the consortium (including the Commission Services)	

An Evaluation of the Use of Data Mining in Counter-Terrorism

Table of Contents

Who Should Read This Paper? Parties that may find this paper of interest include government agencies considering the deployment of data mining technologies in the counter-terrorism context, policy makers in the field of national security, counter-terrorism and law enforcement agencies, bodies that oversee intelligence or national security activities, and non-governmental organizations focussed on the field of human rights or national security.

Executive Summary

1. Comprehensive assessments of the performance of data mining programmes, whether in the counter-terrorism context or other contexts, are generally not publicly available.

2. Despite the lack of evaluative information, some applications of data mining show promise, and one can intuitively assume that data mining is useful in performing traditional investigatory tasks.
3. Where data mining is applied in the counter-terrorism context, however, there may be more acute risks of human rights violations than is the case in other contexts.
4. Limiting the use of counter-terrorism data mining to that of an analytical tool in targeted investigations and to applications that do not rely on personal data would minimize the potential for human rights infringement.
5. More public studies of and research on the performance of data mining programmes providing demonstrable results could help to establish a realistic view of the promise of data mining and open a dialogue on the most sensible manner in which it can be applied in the counter-terrorism context.

Summary of Recommendations

1. Governments should require rigorous testing of data mining products before implementing them for counter-terrorism purposes.
   1.1. Testing should be carried out on simulated data that reflects to the closest degree possible the forms and types of data to which the deployed product is intended to be applied.
   1.2. Testing should be performed by a competent body that is completely independent of the agency or agencies which would use the technology if it were to be deployed, and that body should possess the appropriate expertise in scientific evaluation and assessment methodology.
   1.3. Ideally, results of tests involving deployed technologies should be published for public review so that citizens may assess whether the technology represents a worthwhile investment of public funds.
   1.4. Only those data mining programmes that can demonstrate their effectiveness in the test setting should be allowed to be deployed
2. Following implementation, programmes should be reviewed on a regular basis and monitoring mechanisms updated accordingly.
3. Governments should establish parameters for the types of data that will be subject to data mining programmes or exercises.
   3.1. Parameters should be based on the minimum amount and types of personal data necessary to conduct the analysis for the aims that it seeks to achieve.
   3.2. Where a data mining programme is applied to personal data, use of the programme should be confined to the greatest extent possible to investigatory applications centring on known suspects and endeavour to comply with traditional standards governing government intrusion into the private life of individuals.

1. Introduction

This paper aims to address the potential benefits and costs of data mining programmes as implemented for counter-terrorism purposes. These issues are of interest on at least two levels. First, the effectiveness of data mining programmes is of relevance from the perspective of counter-terrorism agencies. Any efforts aimed at combating terrorism naturally face a limited set of resources. The economic costs involved in building a new

system for data mining can include materiel (hardware, etc.), software, labour in programming, systems engineering, testing, and personnel training. There will also be maintenance costs involved once the system is in place. If the system is not sufficiently effective, however, it will not provide the desired return on investment ("ROI"). In addition to economic costs, there is also the economics of human attention. Thus, the time spent working on producing, evaluating, and investigating the results of data mining efforts can divert human attention—whether at the level of analysts or operative agents—from other areas of interest.

Second, the effectiveness of counter-terrorism data mining programmes is of relevance for assessing the compatibility of such programmes with international human rights standards. Under international human rights law, any government measure that interferes with a human right has to conform to the principle of proportionality. Clearly, if a measure that affects a human right has little to no benefit toward the accomplishment of its aim, it cannot be proportionate. We will address the issue of proportionality in relation to counter-terrorism data mining in more detail in Deliverable D08.3.

In trying to measure effectiveness, we face difficulties on a number of levels. First of all, if we seek to determine a kind of ROI or perform a cost-benefit analysis, some of the costs—such as human attention—are not easily quantifiable. Not to mention the fact, that even straightforward economic costs for specific programmes may be difficult if not impossible to obtain or assess in a comprehensive fashion.[1] Assessing the benefits is equally difficult: Foiled plots are not necessarily always reported. Of those pre-emptive arrests that are publicized, it is generally impossible to know whether data mining played a role and, if so, what role it played exactly. Similarly, comprehensive figures on the success or failure of particular programmes are not available. Public knowledge of government use of data mining programmes—even if details are not revealed—may also have some deterrent effect. As in other contexts, the existence and force of a deterrent effect is difficult to determine.

Additionally, an issue that is of central importance concerns how data mining technologies are implemented and in what contexts. This factor plays a role both in terms of the benefits and costs. As we will see below, certain methods will prove more effective in certain contexts than others. Furthermore, methods that do prove initially successful may require constant adjustments in order to tune them to newly available data—which may also reflect changes in the modus operandi of terrorists. Another consideration is the amount of data on which a data mining operation is performed, not to mention *whose* data is made available to the operation. The former has a direct correlation to the number of "failures" that a data mining programme will yield, the latter has an impact on whom will be implicated by data mining programmes. Targeted approaches would intuitively seem less likely to have an adverse impact on innocents. These considerations remind us that it is impossible to make generalized appraisals of the use of data mining for counter-terrorism purposes in the abstract; rather, any final assessments must be context-specific.

This paper is organized as follows: First, we will discuss what data mining is in Section 2, referring back to our definition and discussion in D08.1. In Section 3, we will look at the

[1] In some instances, the budget figures of counter-terrorism programmes are considered classified information. See, e.g., Office of the Director of National Intelligence, *DNI Releases Budget Figure for FY 2012 Appropriations Requested for the National Intelligence Program*, ODNI News Release No. 4-11 (2011). See also U.S. Government Accountability Office, "Secure Flight Certification", GAO-10-535R (2010) at 6–7 on the difficulties of providing a reliable life-cylcle cost estimate for the Secure Flight programme.

benefits of data mining and examine its track record in other contexts. Turning to the application of data mining in the counter-terrorism context, we examine the aims for which data mining programmes have been proposed in this context. On that basis, we develop a typology of programmes in order to distinguish aspects pertaining to the effectiveness and human impact of different types of programmes. We then look at reports of the performance of data mining programmes in the counter-terrorism context. In Section 4, we turn to the costs connected to the use of data mining in this context. Then, we will seek to weigh these positive and negative aspects in Section 5. Lastly, conclusions and a set of recommendations are offered in Section 6.

2. What is Data Mining?

In our first deliverable, D08.1, we cursorily addressed the issue of defining and describing data mining. There, we touched upon the fact that, even among practitioners in data mining, the term can have different meanings. In D08.1, we provisionally adopted a very broad definition without actually attempting to formulate a formal definition, primarily because we were interested in examining a wide range of activities that had significant implications for human rights but might not fall within narrower definitions. There we articulated our definition as "the use of information technology to attempt to derive useful knowledge from (usually) very large data sets". It is not necessary to articulate a definition here, although we acknowledge criticism that the original definition was too broad.[2] Nonetheless, it is worth pointing out that the fuzziness of data mining as a concept makes any assessment of data mining in the abstract considerably more difficult. In essence, data mining represents a variety of techniques that may be applied (with varying degrees of success) to a variety of different tasks. The types of techniques applied are constantly being experimented with and expanded as the types of tasks for which these techniques are believed to be useful are also further explored. Our initial survey revealed a number of different applications within the counter-terrorism context alone.

Generally, data mining involves the application of one or more algorithms[3] to a set of data to organize that data in a particular way, reveal relationships between data items, and/or assign some value to data items to indicate their significance to a particular query. Well-known applications of data mining include Google's PageRank function which attempts to predict what webpage on the internet a search engine user is most interested in,[4] Amazon's profiling of customers to predict what books they may be interested in, and credit card monitoring which aims to identify potentially fraudulent uses of a client's credit card.

Data mining is often tied to the notion of profiling.[5] Data mining may rely on a profile or model in order to find items within a database that most closely match that profile. This

[2] R. Barquin, "To Data Mine or Not to Data Mine in the Fight Against Terrorism", BeyeNETWORK, 24 August 2010, http://www.b-eye-network.com/channels/1020/view/14227.

[3] For more information on specific algorithms, readers may be interested to consult, e.g., B. Anrig, W. Browne and M. Gasson, "The Role of Algorithms in Profiling", in M. Hildebrandt and S. Gutwirth (eds.), *Profiling the European Citizen* (Springer, 2008). "An alorithm [sic] is like a recipe: given some input it describes the steps to be performed in order to generate the output." Ibid., n. 45, at 65.

[4] However, the core PageRank function is is supplemented by measures aimed at counteracting attempts to game search results. See, e.g., Google's Webmaster Guidelines, available at http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=35769 (version of 12/09/2010) ("The spam reports we receive are used to create scalable algorithms that recognize and block future spam attempts." "...avoid links to web spammers or 'bad neighborhoods' on the web, as your own ranking may be affected adversely by those links.").

[5] In the law enforcement setting, profiling has been defined as "the systematic association of sets of

type of data mining falls within what is termed "supervised" methods. For instance in the area of fraud detection, data analysis will rely on a model based upon past instances of fraud that are known. Unsupervised methods, on the other hand, may more closely resemble the kind of "data dredging" or fishing expedition exercises with which data mining has often been associated in popular accounts. Unsupervised uses of data mining may simply look for any existing patterns in data or for unusual data items—i.e. "outliers".[6] A pattern identified in data may in turn be used to generate a profile. This exercise may be done by applying unsupervised data mining on a data set that contains known items of the type being sought. It is then possible to examine what elements, if any, distinguish those items from other items in the data set.

In many circles, the term "profiling" automatically conjures up negative associations with racial or ethnic profiling. It is important to note that profiling need not involve consideration of any personal attributes whatsoever. An online bookstore, for instance, may create profiles of its customers' book-purchasing habits—such as what genres tend to be purchased, at what times purchases tend to be made, and how much money tends to be spent per purchase—without collecting information on the gender, race, national origin, or religious affiliation of those customers. Nonetheless, there is continued interest in using characteristics such as race, national origin, and religion as a basis for security-related measures.[7] Additionally, automated profiling tools currently applied in the counter-terrorism context within some Western states likely include consideration of at least nationality.[8] However, it is also important to note that data mining is not synonymous with profiling. Data mining may be used to conduct profiling exercises, but it may also be applied in other manners.

The distinction between supervised and unsupervised methods reveals that data mining may be applied in either a directed or undirected manner. In other words, data mining may be part of a highly designed approach in which the user generally knows what kind of information he or she is looking for. The data mining application is then designed specifically in an attempt to reveal that kind of information. On the other hand, data mining may also be used in an exploratory fashion by simply applying stock algorithms to data sets to see what results are produced. This sort of undirected use of data mining always carries with it the risk that the results it produces will be obvious, uninteresting, or irrelevant.[9]

3. Benefits

---

physical, behavioural or psychological characteristics with particular offences and their use as a basis for making law-enforcement decisions." M. Scheinin, "Implementation of General Assembly Resolution 60/251 of 15 March 2006 Entitled "Human Rights Council": Report of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism", A/HRC/4/26 (2007) at 33.

[6] See R. J. Bolton and D. J. Hand, "Statistical Fraud Detection: A Review" (2002) 17:3, Statistical Science, 235–55 at 236.

[7] See, e.g., "Our view on airport screening: Why Israel's air security model wouldn't work in the USA", USA Today, 21 December 2010, http://www.usatoday.com/news/opinion/editorials/2010-12-22-editorial22_ST_N.htm.

[8] Following the Christmas Day bomb attempt on a Detroit-bound flight, several Western nations voiced the opinion that certain nationals should be subjected to higher security scrutiny, suggesting that nationality represented a factor for risk assessment. See DETECTER Deliverable D06.3, n. 154 and accompanying text.

[9] Cf. M. DeRosa, *Data Mining and Data Analysis for Counterterrorism* (26 June 2009), p. 3.

As a form of IT-supported data analysis, data mining promises to provide similar benefits to that of other forms of IT-based data analysis. Perhaps first and foremost is the benefit of processing speed. Today's capabilities in terms of digital storage capacities and processing speeds make it possible to store and retrieve large amounts of information and perform calculations much faster than human beings and without human error. As one example of modern computing power, the FBI claimed that the deployment of the technology behind its Investigative Data Warehouse had reduced the time to perform certain tasks from 32,000 hours to half an hour.[10]

The use of data mining has traditionally been connected with investigating very large amounts of data. As businesses of all kinds began to accumulate databases of information pertaining to sales, customer relations, spending, etc., data mining promised a method for analyzing these growing bodies of data so that it could serve to better understand and improve business processes. The problem, of course, was not limited to the business world. In a seminal 1996 article, Fayyad, Piatetsky-Shapiro, and Smyth referenced the increasing size and depth of databases in the realms of medicine and the astronomical sciences in particular.[11] They indicated that, even at that time, databases containing millions of items were "becoming increasingly common" and each item might have as many as hundreds or thousands of different fields of information associated with it.[12]

Related to this promise of providing analysis of large data sets is the emerging ability of data mining programmes to provide graphical representations of trends, patterns in or connections among data. In this way, it is hoped that these special visualization tools can support analysis by providing a quick picture of informational relevance.

One aspect for which data mining is renowned and that distinguishes it from other forms of data processing, however, is its ability to uncover relationships or patterns within data that investigators may not even think to inquire after. A much-cited example from the field of market basket analysis was the discovery that, at one particular store, beer and diapers were often purchased together within a certain span of hours.[13]

3.1. Performance in Other Contexts

A number of data mining success stories have been reported in contexts other than counter-terrorism:

- Marketing

  Credit card company American Express reportedly saw a 10-15% increase in credit card use after implementing a system for purchase analysis.[14]

- Credit Card Fraud

[10] Chiliad. (2006). *Chiliad Success Story: Federal Bureau of Investigation.* Retrieved November 23, 2009, from http://www.chiliad.com/docs/ChiliadCaseStudy_FBI.pdf, p. 3.
[11] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "From Data Mining to Knowledge Discovery in Databases" (1996) 17(3), AI Magazine, 37–54 at 38.
[12] Ibid., p. 38.
[13] See M. Whitehorn, "The parable of the beer and diapers", The Register, 15 August 2006, http://www.theregister.co.uk/2006/08/15/beer_diapers/.
[14] Ibid., p. 38.

Amy Belasco of the Congressional Research Service examined a system described in a 1999 article that was designed to detect credit card fraud. She observed that when the developers combined different algorithms they were able to detect up to 50% of fraudulent transactions with the system.[15] In 2003, it was reported that credit card issuers U.S. Bancorp and Wachovia had reduced incidents of fraudulent credit card use by 70% since 1992.[16] At that time, credit union insurer Credit Union National Association was reportedly including the use of such fraud detection software as one requirement for credit unions to qualify for insurance.[17]

- Telecommunications

In the late 90s, U.S. telecommunications provider AT&T faced a billing scam in which customers who visited a Moldovan pornography website had software surreptitiously installed on their computers. The software would disconnect the connection to AT&T and dial an expensive toll number in Moldova. By mining its database of call data, AT&T together with the Moldovan authorities was able to locate the actors behind the scheme.[18]

- Medicine

Drug-producer Novartis reportedly used data mining tools to analyze cases of infant leukaemia. Its findings suggested that instances of the disease in infants could be grouped into three different categories, providing an indication that three different types of treatment might be called for.[19]

- Law Enforcement

Joyce Knowlton, an investigator at the Stillwater State Correctional Facility in Minnesota used i2, Inc.'s Analyst Notebook product to uncover a drug smuggling ring that was operating within the prison. Knowlton entered prisoner call record data into the software, which revealed a pattern of calls between prisoners and a recent parolee. By comparing the call patterns with prisoner financial records, she was able to determine a pattern of money flows. On this basis, she began to monitor the telephone conversations of certain inmates and ascertained that they were using coded messages in connection with the drug smuggling activities.[20]

---

[15] A. Belasco, "Total Information Awareness Programs: Funding, Composition, and Oversight Issues", RL31786 (March 21, 2003) at CRS-16.. There were of course also a significant number of false positives. We will discuss this aspect further below. The system she refers to is described in Philip K. Chan, Wei Fan, Andreas L. Prodromidis, and Salvatore J. Stolfo (1999) Distributed Data Mining in Credit Card Fraud Detection, available at http://cs.fit.edu/~pkc/papers/ieee-is99.pdf.

[16] O. Port, 'Smart Tools: Companies in health care, finance, and retailing are using artificial-intelligence systems to filter huge amounts of data and identify suspicious transactions' (Spring 2003), *Business Week*, available at http://www.businessweek.com/bw50/content/mar2003/a3826072.htm.

[17] Ibid.

[18] J. Markoff, "Taking Spying to Higher Level, Agencies Look for More Ways to Mine Data", New York T imes, 25 February 2006, http://www.nytimes.com/2006/02/25/technology/25data.html?ei=5088.

[19] O. Port, 'Smart Tools: Companies in health care, finance, and retailing are using artificial-intelligence systems to filter huge amounts of data and identify suspicious transactions' (Spring 2003), *Business Week*, available at http://www.businessweek.com/bw50/content/mar2003/a3826072.htm.

[20] Ibid.

The analysis of PNR and other traveller data, such as "Advanced Passenger Information" data, has also reportedly been instrumental in ferreting out international drug smuggling. In the case of *US v. McKenzie*,[21] for example, the defendant was approached by federal drug enforcement officials solely on the basis of the fact that his PNR data fit the drug courier profile. The defendant in *Lewis v. Texas*[22] was referred to local Customs agents at Dallas/ Fort Worth International Airport on the basis of his travel patterns, the fact that he paid cash for his flight, and was believed to be related to another Lewis who had been arrested for transporting heroin into the US.[23] The EU Commission has also cited examples where human traffickers were uncovered by linking them to false documents and where drug smugglers were exposed on the basis of their use of stolen credit card numbers to purchase flights.[24]

Law enforcement officials in eastern Virginia used data mining to identify when and where complaints about random gunfire were most likely to occur around the time of the New Year's Eve holiday. The analysis was used to plan police deployment for the 2003-2004 holiday. The initiative reportedly saw a 47% decrease in random gunfire complaints on New Year's Eve and a 26% decrease over the course of the two-day holiday.[25] The number of guns seized over the holiday period also increased from 13 in the previous year to 45 within the course of the initiative.[26] These gains were registered despite the fact that fewer police were deployed over the holiday period as part of the initiative's targeted strategy. As a result, the initiative saw savings of 15,000 USD in personnel costs.[27]

- Anti-Money Laundering

The field of Anti-Money Laundering (AML) overlaps with that of counterterrorism since AML systems are also used to identify terrorist financing. In 1990, the US Department of the Treasury established the Financial Crimes Enforcement Network (FinCEN) as a support unit for state and federal law enforcement. Its purpose was to provide information and analysis and identify targets for investigation in the field of money laundering and financial crimes.[28] FinCEN employs an automated analysis system known as the FinCEN Artificial Intelligence System (FAIS) for these purposes.[29] In a 1995 report, the Congressional Office for Technology Assessment wrote that the system had had "some clear successes" but the true extent to which FAIS had demonstrated its usefulness was unclear.[30] In 2003, however, the General

[21] No. CR 08-1669, slip op., 2010 WL 1795173 (D. N.M. April 13, 2010).
[22] No. 05-00-01204-CR, 2001 WL 946818 (Tex. App. Aug. 22, 2001).
[23] It was later determined, however, that he bore no relation to the other Lewis in question. Ibid., note 2.
[24] European Commission, "Commission Staff Working Paper: Impact Assessment", SEC(2011) 132 at 12.
[25] C. McCue, "Data Mining and Predictive Analytics: Battlespace Awareness for the War on Terrorism" (2005) 13:1&2, Defense Intelligence Journal, 47–63 at 264.
[26] Ibid., p. 264.
[27] Ibid., p. 264.
[28] Office of Technology Assessment, U.S. Congress, "Information Technologies for Control of Money Laundering", OTA-ITC-630 (September 1995) at 43.
[29] Ibid., p. 44.
[30] Ibid., p. 46. The report states: "In spite of some clear successes, evaluation of FinCEN's help to law enforcers is difficult. FinCEN itself has little direct feedback from clients and thus little knowledge of the results of its referrals. Some field level law enforcement agents are skeptical; some told OTA that they have not been aware of any assistance from the agency. IRS, Customs, DEA, and FBI agents who

> Counsel of the Department of the Treasury indicated that FAIS had supported 2,692 terrorism investigations and had proactively generated 519 case referrals to law enforcement since 11 September 2001.[31]

These anecdotes represent just a select few examples. Colleen McCue[32] and Christopher Westphal[33] provide other illustrations of how data mining may be applied in the law enforcement and security contexts. "Case studies" and "client testimonials" can also generally be found on the websites of any data mining technology provider.[34] Hard numbers with regard to outcomes or quantifiable improvements, however, are rarely cited and are almost never provided in extensive detail.

3.2. Aims of Data Mining in the Counter-Terrorism Context

As with enterprise management as noted above, intelligence and investigatory work in the field of counter-terrorism also faces the problem of information overload.[35] One of the primary themes of the findings of the 9/11 Commission was that information pertaining to the unfolding of the 9/11 plot was available within intelligence files and databases, but that analysts had failed to put all the information together.[36] Thus, in the same way that it was able to assist businesses in analyzing internal information, data mining seemed to offer the ability to bring such intelligence information together quickly and automatically to the benefit of overwhelmed analysts.

In at least some instances, the use of data mining in intelligence and law enforcement merely represents the application of information technology to the same tasks that such agencies have traditionally performed in the past.[37] Thus, data mining may allow these agencies to work in a much faster, more efficient, and perhaps more organized manner than in the past. The ability of data mining to reveal associations that analysts might not think to inquire after may have also offered some hope that data mining would not only assist in

---

have worked "on the street" or mounted active operations told OTA that they relied much more heavily on their own agencies' intelligence units, on undercover agents, or on tips from informants. However, there may be reasons for this; leads generated by FinCEN may be passed through higher levels of a user agency to its agents without being identified as to source. FinCEN information may be discounted or ignored by some agents who are not used to dealing with that kind of data. Some agents who talked with OTA had not been on the street for several years, and FinCEN's most sophisticated products have been introduced in the last year or two. Higher level comments may well be intended to protect an agency's own image and budget." Ibid.

[31] Counterterror Initiatives in the Terror Finance Program: Hearing Before the Senate Committee on Banking, Housing, and Urban Affairs, 108th Cong. (2003) (Written Testimony of David D. Aufhauser, General Counsel, U.S. Department of the Treasury).

[32] C. McCue, *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis* (Amsterdam: Elsevier/Butterworth-Heinemann, 2007).

[33] C. Westphal, *Data mining for intelligence, fraud & criminal detection: Advanced analytics & information sharing technologies* (Boca Raton , FL: CRC Press, 2009).

[34] See, e.g., http://www.acl.com/customers/success_stories.aspx; http://www.spss.com/success/.

[35] See, e.g., McCue, supra note 25, 48, 50.; McCue, supra note 32, p. 32.

[36] National Commission on Terrorist Attacks upon the United States, *The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks upon the United States*, (Washington, 2004).

[37] See K. A. Taipale, "Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data" (2003 / 2004), Columbia Science and Technology Law Review, 2 at 23–25.

performing traditional investigation tasks but could uncover connections or leads that traditional techniques would not.[38]

Based on our findings in D08.1, there are a number of objectives that law enforcement, security agencies, and intelligence agencies hope to accomplish through data mining. One such objective concerns the discovery of terrorists and terrorist networks. Both of the two most renowned programmes in the US that came to be strongly associated with the words "data mining"—TIA[39] and MATRIX[40]—sought, at least in part, to perform this function. Another of the most renowned programmes in this context, the German *Rasterfahndung*,[41] also had this objective. Intelligent flight screening measures, such as those involved in the US Automated Targeting System[42] and Secure Flight,[43] represent yet another form of data mining which seeks to identify terrorists in real time.

A second objective that is sought through the use of data mining is the generation of profiles—in the flight screening context, based on travel or other behavioural patterns that are then applied to travellers.[44]

A third objective which counter-terrorist data mining seeks to achieve is the assessment of risks. This might take the form of threat detection systems, such as automated analysis or specialized filtering of communications—for example the NSA's programme[45]—or the automated analysis of open source materials such as news articles in order to predict when and where attacks will take place and ideally by whom they are being orchestrated.[46] Threat detection may also involve the analysis of video data to detect emerging threats in real time or simply identify "suspicious" behaviour. The FP7 project SAMURAI[47] includes this kind of application of data mining to video data to detect "abnormal behaviour" as it is recorded by security cameras. Additionally risk assessment may be geographical, using maps of criminal or other activity of interest to predict where future terrorist activity may occur.

A fourth objective which the use of data mining seeks to achieve is the provision of analytic assistance, for instance through the automated prioritization of information[48] or through

---

[38] Hence, the preoccupation with the discovery of "novel intelligence". See, e.g., NSA's "Novel Intelligence From Massive Data (NIMD)" programme, D08.1, § 2.2.6.
[39] See D08.1, § 2.2.4.
[40] See M. Shnayerson, 'The Net's Master Data-miner' (December 2004), *Vanity Fair*, available at http://www.vanityfair.com/ontheweb/features/2004/12/matrix200412?p; see also n. 52 and accompanying text below. For additional information, see D08.1, § 2.2.5. Plans for the MATRIX may have gravitated more towards providing a general law enforcement tool, including more enhanced search features, by the time it began to take shape as a multi-state project.
[41] See D08.1, § 2.4.1.
[42] See D08.1, § 2.2.3.
[43] See D08.1, § 2.2.2. Ultimately, the performance of a risk profile assessment was not included in the deployed version of Secure Flight. Currently, the programme only performs watchlist matching and entity resolution functions. Stephen M. Lord, Director of Homeland Security and Justice Issues to James Thurman, 20 January 2011, 'Secure Flight questions', E-mail.
[44] European Commission, "Communication from the Commission: On the global approach to transfers of Passenger Name Record (PNR) data to third countries", COM(2010) 492 final (2010) at 4; see also M. DeRosa, *Data Mining and Data Analysis for Counterterrorism* (March 2004), p. 11.
[45] See e.g., Complaint, Hepting v. AT&T, Corp., (N.D. Cal. Jan. 31, 2006), available at http://www.eff.org/files/filenode/att/att-complaint.pdf.
[46] Recorded Future represents one example of this pursuit. See http://www.analysisintelligence.com/.
[47] http://www.samurai-eu.org/Home.htm.
[48] This function could take a number of forms. One example are operations typically performed by some varieties of search engines, where search results are ranked according to some function of "relevance". Another form are automated alerts which rely on monitoring processes that look for a

visual representations of data. Examples are the Genoa II project within TIA,[49] NSA's NIMD programme,[50] and the Defense Intelligence Agency's Insight Smart Discovery.[51]

On this basis, we have established a typology of four applications of data mining in the counter-terrorism context:

1. Discovery of terrorists and terrorist networks
2. Profile Generation
3. Risk Assessment
4. Analytic Aids

This typology is not intended to be absolutely comprehensive or exclusive. As noted above, new techniques and applications are being developed all the time. Therefore, even if this list represents an accurate reflection of current data mining applications in the counter-terrorism context, additional types of applications may be developed in the future.

3.3. Performance in Counter-Terrorism

As we noted above, it is difficult to find evidence of instances where data mining programmes provided the key toward the foiling of a terrorist plot, the arrest of a terrorist or other intervention on the part of authorities. Perhaps the best figures are available for the German *Rasterfahndung* since it was subjected to judicial review and the court's reasoning is recorded in a published opinion. That programme appears to have been more or less a complete failure. It must be pointed out, however, that the *Rasterfahndung* appears to represent a rather crude form of data mining in which information technology probably only played a very minor role. Nonetheless, there is evidence that US authorities followed the same line of thinking, and the State-based programme, the MATRIX, may have begun with a similar design in mind.[52] Reports of other similarly crude forms of investigation have also been reported in US media.[53] That a data mining programme is linked to a particular instance of success in the media or otherwise in a public source is extremely rare.

Below, we outline some of the instances in which information concerning the performance of a specific programme has been revealed in publicly available sources.

---

particular level of correspondance with an established profile.

[49] See D08.1, § 2.2.4.

[50] See D08.1, § 2.2.6.

[51] D08.1, § 2.2.9.

[52] Reportedly, the idea behind the MATRIX began when Hank Asher tried to design algorithms to isolate the 9/11 hijackers within a number of databases. His programme ultimately turned up 419 individuals who were assigned particularly high suspicion scores on the basis of the profile Asher had come up with. Asher's profile evidently assumed that other terrorists would have the same sort of characteristics as the 9/11 hijackers or at least those characteristics which Asher believed the hijackers would have. Asher's profile, however, was more sophisticated than that used in the *Rasterfahndung* as it apparently included factors such as voter registration and length of US residence. See M. Shnayerson, 'The Net's Master Data-miner' (December 2004), *Vanity Fair*, available at http://www.vanityfair.com/ontheweb/features/2004/12/matrix200412?p. The generation of suspicion scores, however, was reportedly dropped from the initiative by 2005. H. C. Relyea and J. W. Seifert, "Information Sharing for Homeland Security: A Brief Overview", RL32597 (2005) at 12.

[53] One example is the report that the FBI at one time sought to identify terrorist suspects by identifying purchases of typical Middle Eastern grocery items within supermarket purchase data from LA-area markets. See J. Stein, 'FBI Hoped to Follow Falafel Trail to Iranian Terrorists Here' (2 November 2007), *CQPolitics*, http://www.cqpolitics.com/wmspage.cfm?docID=hsnews-000002620892.

Rasterfahndung

According to publicly available sources, the *Rasterfahndung* ultimately involved trawling through the data of some 8.3 million people—over 10 percent of the German population.[54] The *Rasterfahndung* involved a profile-based search through various databases and data sets. The profile was likely based on the individuals involved in the 9/11 hijacking[55] and was generated by a coordinating group composed of officials from the various states (*Länder*) of Germany that was headed by the Federal Criminal Police Office and featured representatives from Federal Border Control, the Federal Office for the Protection of the Constitution, and the Federal Intelligence Agency.[56] The profile consisted of the following characteristics: male, between 18 and 40 years of age, student or former student, Muslim, with place of birth in or nationality of one of certain specific countries that had predominantly Muslim populaces.[57] Among the databases that were searched were university databases, residential registration offices,[58] and the national alien registry.[59] The search conducted by the local agencies reportedly identified 31,988 "records" that were reported to the Federal Criminal Police Office.[60] The data that was reported in these records varied from state to state and the particular source from which it was acquired. In North Rhine-Westphalia, for instance, data delivered from residential registration offices included: first name and surname, name at birth, place of birth, nationality, address, any secondary residences, religion, marital status, children, the tax office that handled the individual's tax statements, date on which residence was established, date on which the individual moved out of the registration office's jurisdiction.[61] Records from universities and other institutions of higher education included the individual's field of study.[62]

In at least one case, it was later discovered that some of the individuals named in these records had been mistakenly included. The state of North Rhine-Westphalia, for instance, subsequently determined that a total of 1,187 of the individuals whose data had been passed on to the federal authorities did not actually match all the profile characteristics.[63] Data from these records was cross-referenced with other data sets, which reportedly included files concerning the holders of pilot's licenses and individuals who were subjected to background checks for handling radioactive substances.[64] Whenever the Federal Criminal Police Office determined that someone within the cross-referenced data sets matched an individual within the set of 31,988 records, that individual was put into a "results" file that was made available to the state criminal agencies.[65] Reportedly, the results file ultimately held information on 1,689 individuals who were subject to closer investigation by German

---

[54] T. Berndt and D. Schiffermüller, *Rastern für Europa – Minister Schily und sein Fahndungsflop*, Panorama, NDR (Hamburg), 08 April 2004, http://daserste.ndr.de/panorama/archiv/2004/erste8498.html.
[55] National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment* (Washington, D.C: National Academies Press, 2008), p. 215, Box H.3.
[56] *Rasterfahndung,* 1 BvR 518/02 (Bundesverfassungsgericht, 04 April 2006), para. 8.
[57] Ibid., para. 8.
[58] Anyone who resides in Germany has to register at these local offices so that everyone residing within the country has certain personal data on file that is linked with a local address.
[59] *Rasterfahndung,* 1 BvR 518/02, para. 8.
[60] Ibid., para. 9.
[61] See Ibid., para. 13.
[62] Ibid., para. 13.
[63] Ibid., para. 30.
[64] Ibid., para. 9.
[65] Ibid., para. 9.

police.[66] Ultimately, the *Rasterfahndung* did not result in even a single criminal indictment for terrorism-related offences.[67]

CAPPS II

CAPPS II was a proposed airline passenger pre-screening programme that was intended to be administered by the US Transportation Security Administration (TSA) and replace the CAPPS I screening that was being carried out by commercial airlines.[68] After data mining programmes such as TIA and CAPPS II had gained significant media attention, the US Congress mandated in 2003 that the Government Accountability Office (GAO)[69] investigate CAPPS II and ensure that it met eight criteria.[70] One criterion required the TSA to demonstrate that it "has stress-tested and demonstrated the efficacy and accuracy of all search tools in CAPPS II and has demonstrated that CAPPS II can make an accurate predictive assessment of those passengers who may constitute a threat to aviation".[71] In its 2004 Report on CAPPS II, the GAO indicated that several goals had not yet been met, including that TSA had not yet "stress tested and demonstrated the accuracy and effectiveness of all search tools to be used by CAPPS II".[72] Notably, this statement includes no mention of whether the demonstration of the threat assessment aspects of CAPPS II were still outstanding. It is thus unclear whether this omission indicates that this objective was met or was still considered to be a part of the other testing of system performance which was yet to be done.

The development of CAPPS II was cancelled in 2004. It is unknown whether any concerns about the proposed system's effectiveness that may have existed played any role in the abandonment of the project. In a 2005 report, the GAO indicates that the programme was discontinued due to "a variety of delays and challenges".[73] Currently, the GAO website heralds as one of its recent accomplishments "[p]rompting the saving of over $300 million

[66] National Research Council, supra note 55, p. 215., Box H.3.
[67] *Rasterfahndung*,, para. 10.
[68] See D08.1, § 2.2.1 for additional details.
[69] Then still known as the "General Accounting Office".
[70] The eight criteria were: "(1) a system of due process exists whereby aviation passengers determined to pose a threat and either delayed or prohibited from boarding their scheduled flights by the TSA may appeal such decision and correct erroneous information contained in CAPPS II; (2) the underlying error rate of the government and private data bases that will be used both to establish identity and assign a risk level to a passenger will not produce a large number of false positives that will result in a significant number of passengers being treated mistakenly or security resources being diverted; (3) the TSA has stress-tested and demonstrated the efficacy and accuracy of all search tools in CAPPS II and has demonstrated that CAPPS II can make an accurate predictive assessment of those passengers who may constitute a threat to aviation; (4) the Secretary of Homeland Security has established an internal oversight board to monitor the manner in which CAPPS II is being developed and prepared; (5) the TSA has built in sufficient operational safeguards to reduce the opportunities for abuse; (6) substantial security measures are in place to protect CAPPS II from unauthorized access by hackers or other intruders; (7) the TSA has adopted policies establishing effective oversight of the use and operation of the system; and (8) there are no specific privacy concerns with the technological architecture of the system." Department of Homeland Security Appropriations Act, 2004, Pub. L. No. 108-90, § 519, 117 Stat. 1137, 1155-56 (2003).
[71] Ibid., § 519(a)(3).
[72] U.S. General Accounting Office, "Aviation Security: Computer-Assisted Passenger Prescreening System Faces Significant Implementation Challenges", GAO-04-385 (2004), http://www.gao.gov/new.items/d04385.pdf, at 4.
[73] U.S. Government Accountability Office, "Aviation Security: Secure Flight Development and Testing Under Way, but Risks Should Be Managed as System Is Further Developed", GAO-05-356 (2005) at 2. See also J. W. Seifert, "Data Mining and Homeland Security: An Overview" (April 3, 2008) at 9–11.

from the cancellation of the Transportation Security Administration's CAPPS II Program."[74] This statement seems to suggest that investment in the CAPPS II project would have been in some sense wasteful.[75] In its 2008 report on data mining and privacy, the US National Research Council noted "data mining for credit scoring is widely acknowledged as an extremely successful application of data mining, while the various no-fly programs (e.g., CAPPS II) have been severely criticized for their high rate of false positives."[76]

Secure Flight

Secure Flight is the flight screening system currently in place in the United States.[77] Once it emerged that the TSA had begun work on Secure Flight in lieu of CAPPS II, Congress renewed the same set of requirements the following year for "CAPPS II or Secure Flight or other follow on/successor programs", adding two additional requirements.[78] The mandate was then renewed for each successive year through 2009.[79] Accordingly, the GAO attempted to evaluate Secure Flight and verify that it fulfilled the ten conditions over the course of that time period. Two of those conditions related directly to the issue of Secure Flight's effectiveness. The first concerned an assessment of the extent of false positives, and the other concerned the performance of "stress testing" and testing of the efficacy and accuracy of Secure Flights "search tools".[80] The GAO eventually reported that nine of the ten conditions had been "generally achieved" in April 2009, which included both conditions related to technical effectiveness.[81] It is unclear from the GAO 2009 Report, however, to what extent the GAO was directly involved in testing the efficacy of the system. A GAO official has confirmed that employees of the GAO observed tests of the system.[82] Yet, it appears that the GAO also relied on statements from TSA officials that declared that the system was performing up to established parameters. For instance the GAO reports that "TSA officials stated that they tested the system's false-positive performance during Secure Flight's parallel testing with selected air carriers in January 2009 and found that the false-positive rate was consistent with the established target and program's goals."[83]

---

[74] http://www.gao.gov/careers/hsj.html.
[75] It is also possible that the "abandonment" of CAPPS II merely represented a kind of "re-branding" of the project accompanied by the decision to change certain aspects of it.
[76] National Research Council, supra note 55, p. 214. We discuss the issues of false positives and false negatives in Section 3.5.2.
[77] See D08.1, § 2.2.2.
[78] Department of Homeland Security Appropriations Act, 2005, Pub. L. No. 108-334, § 522, 118 Stat. 1298, 1319-20 (2004) The two additional requirements were "(9) the TSA has, pursuant to the requirements of section 44903 (i)(2)(A) of title 49, United States Code, modified CAPPS II or Secure Flight or other follow on/successor programs with respect to intrastate transportation to accommodate States with unique air transportation needs and passengers who might otherwise regularly trigger primary selectee status; and (10) appropriate life-cycle cost estimates, and expenditure and program plans exist."
[79] Department of Homeland Security Appropriations Act, 2006, Pub. L. No. 109-90, § 518(a), 119 Stat. 2064, 2085 (2005); Department of Homeland Security Appropriations Act, 2007, Pub. L. No. 109-295, § 514(a), 120 Stat. 1355, 1379 (2006); Consolidated Appropriations Act, 2008, Pub. L. No. 110-161, div. E, § 513(a), 121 Stat. 1844, 2072 (2007); Department of Homeland Security Appropriations Act, 2009, Pub. L. No. 110-329, div. D, § 512(a), 122 Stat. 3652, 3682 (2008).
[80] U.S. Government Accountability Office, "Aviation Security: TSA Has Completed Key Activities Associated with Implementing Secure Flight, but Additional Actions Are Needed to Mitigate Risks", GAO-09-292 (2009) at 2., Table 1.
[81] Ibid., p. 9.
[82] Stephen M. Lord, Director of Homeland Security and Justice Issues to James Thurman, 20 January 2011, 'Secure Flight questions', E-mail (on file with authors).
[83] U.S. Government Accountability Office, "Aviation Security: TSA Has Completed Key Activities Associated with Implementing Secure Flight, but Additional Actions Are Needed to Mitigate Risks",

Significantly, neither the GAO nor Congress set the "established target" for false positives,[84] suggesting that this benchmark was internally established.

With respect to the third condition, the GAO went into further detail concerning testing carried out by the TSA. According to the report, the TSA conducted a series of tests of the watchlist matching capabilities of Search Flight, enlisting the support of a third party contractor who was an expert in watchlist matching.[85] The contractor generated a simulated watchlist and a simulated passenger list using software and relying on the expertise of analysts and linguists.[86] The passenger list consisted of about 12,000 records, of which approximately 1,500 represented intended matches to records on the simulated watchlist.[87] In these tests, the Secure Flight system did not identify all the matches that the contractor had.[88] In other words, the system had generated a higher number of false negatives. This result was attributed to the fact that the contractor had used a wider date range for matching birth dates than the Secure Flight system did. However, officials from TSA's Office of Intelligence reviewed the test results and determined that the additional false negatives "did not pose an unacceptable risk to aviation security."[89] It was additionally thought that adding flexibility to the date matching in order to reduce the number of false negatives would also raise the rate of false positives unacceptably.[90] The GAO report makes no mention of false positives encountered in the testing. It is unclear whether this fact is an indication that no false positives were generated in the tests. In testimony from 2005, Paul Rosenzweig, then acting Chairman of the Department of Homeland Security's Data Privacy and Integrity Advisory Committee, suggested that watchlist matching under the old system operated by airline carriers had a match rate of roughly 2%, meaning that on average 35,000 travellers in the US would be flagged for additional scrutiny each day. Rosenzweig indicated that the Secure Flight system promised to bring that rate down to 0.8% (14,000 travellers per day on average).[91]

In addition to testing accuracy, TSA was also required to perform "stress testing" and "performance testing" of Secure Flight. The latter was to assess that the system would still be able to function properly under the levels of data flow and varying conditions that might be experienced during day-to-day operation.[92] Stress testing, on the other hand, was to assess the system's ability to handle rate volumes of data beyond the performance requirements that had been defined for the system.[93] Although the GAO was not satisfied in January 2009 that the testing that had been performed up to that time was adequate to fulfil the condition imposed by Congress, by May of that year, it reported that TSA had "completed performance testing and significantly stress tested the vetting system portion of Secure Flight."[94] It further reported that stress testing established that the vetting system

---

GAO-09-292 (2009) at 13.

[84] Stephen M. Lord, Director of Homeland Security and Justice Issues to James Thurman, 20 January 2011, 'Secure Flight questions', E-mail (on file with authors).

[85] U.S. Government Accountability Office, supra note 83, 14.

[86] Ibid., p. 43.

[87] Ibid., p. 43.

[88] Ibid., p. 14.

[89] Ibid., p. 14.

[90] Ibid., p. 14.

[91] Improving Pre-Screening of Aviation Passengers Against Terrorist and Other Watch Lists: Hearing Before the United States House of Representatives Committee on Homeland Security Subcommittee on Economic Security, Infrastructure Protection and Cybersecurity, 109th Cong. (2005) (Testimony of Paul Rosenzweig, Senior Legal Research Fellow, Center for Legal and Judicial Studies).

[92] U.S. Government Accountability Office, supra note 83, p. 16.

[93] Ibid., p. 16.

[94] Ibid., p. 17.

could successfully perform beyond the defined parameter of 10 names in 4 seconds.[95] The Secure Flight programme received GAO certification on 5 April 2010 once the GAO was satisfied that the final condition concerning life-cycle cost estimates and programme plans had been generally achieved.[96] The system began initial operation on 27 January 2009 on a portion of US flights with a view to gradually extend operations to include all US domestic flights and international flights departing to or from the US.[97]

Automated Targeting System

The Automated Targeting System (ATS) is a screening system employed by US Customs and Border Protection which operates in a similar fashion to the original CAPPS system.[98] News reports have credited the ATS with identifying Raed al-Banna as a potential terrorist and preventing him from entering the United States in 2003. The New York Times, for instance, suggests that ATS registered al-Banna as having '"multiple terrorist risk factors"'.[99] As a result, al-Banna was subjected to questioning upon landing at O'Hare airport and ultimately denied entry into the US. Less than two years later, al-Banna's fingerprints were allegedly matched to the hand of a suicide bomber who detonated a car bomb in Iraq.[100]

ADVISE

ADVISE was a national security programme that was designed to mine data from various databases and provide results in the form of visual analysis and provide suspicion alerts.[101] We have not found any information pertaining to the effectiveness or accuracy of the programme's performance; however, the DHS discontinued the programme in 2007, citing its high maintenance cost and the availability of less expensive commercial off-the-shelf products which could perform the same or similar tasks.[102]

Able Danger

Able Danger was a data mining project of the US Army's Land Information Warfare Agency that was carried out from 1999-2000.[103] Their work had included analysis of the al Qaeda network.[104] Following the September 11 attacks, several individuals involved with the Able Danger project indicated that their work had identified four of the individuals who participated in the plot.[105] Specifically, Able Danger participants suggested that they had

---

[95] Ibid., p. 17.
[96] Ibid.
[97] Ibid., p. 9
[98] See D08.1, § 2.2.3 for additional information.
[99] S. Shane and L. Bergman, "Contained?: Adding Up the Ounces of Prevention", New York Times, September 10, 2006.
[100] Ibid. See also P. Rosenzweig, "Privacy and Counter-Terrorism: The Pervasiveness of Data" (2010) 42, Case Western Reserve Journal of International Law, 625–46 at 634–35.
[101] See D08.1, § 2.2.25 for additional information.
[102] M. J. Sniffen, "DHS Ends Criticized Data-Mining Program", Washington Post, September 5, 2007, http://www.washingtonpost.com/wp-dyn/content/article/2007/09/05/AR2007090500795.html.
[103] See D08.1, 2.2.26 for additional information.
[104] Able Danger and Intelligence Information Sharing: Hearing Before the Senate Committee on the Judiciary, 109th Cong. (2005) (Testimony of Erik Kleinsmith, former Army Major and Chief Intelligence of the Land Information Warfare Analysis LIWA, Lockheed Martin).
[105] Able Danger and Intelligence Information Sharing: Hearing Before the Senate Committee on the Judiciary, 109th Cong. (2005) (Testimony of Mark Zaid, Attorney at Law).

identified Mohammed Atta "and had him linked through associational activities to the blind Sheik and others operating in or around Brooklyn, New York."[106]

Summary

This brief survey of public information relating to the performance of data mining in the counter-terrorism context reveals little in the way of hard statistical information. The most extensive figures available pertain to the German terrorist *Rasterfahndung*. These figures are not at all promising, but given that the exercise essentially consists of the application of an ethnic profile, the poor results are not a great surprise. In the realm of US airline passenger screening, Rosenzweig has provided some potential figures for the number of positives these systems return.[107] However, we do not know how reliable these are or how many of these "matches" represent legitimate positives that indicate that the system is performing as intended. The saga of CAPPS II/Secure Flight, however, is quite telling. Once the programme came under scrutiny, the original plans were scrapped. Then, the successor programme, Secure Flight, which according to initial plans would require two years for development, ultimately required over six years once a number of conditions were imposed. Yet, despite the certification of the Government Accountability Office that the conditions imposed by the US Congress were met, no figures reflecting the performance of Secure Flight such as the percentage of false positives and false negatives have been made public. Isolated anecdotes of "success stories" may be found in media sources, but this handful of purported evidence cannot substitute for comprehensive performance data.

3.4. Problems Relating to Counter-Terrorism Data Mining

There are at least four significant issues that create difficulties for the use of data mining in the counter-terrorism context. The first is the relative rarity with which terrorist activity takes place. Two further issues are those of false positives and false negatives, which represent ways in which data mining fails the task it is supposed to achieve. Lastly, there is the problem of poor data quality which can contribute to the failure of data mining systems to perform as intended.

3.4.1. The Problem of Rarity

In comparison to other events for which data mining has been used in the commercial context, terrorist acts are relatively rare in the Western world. According to the RAND Database of Worldwide Terrorism Incidents, 2006 saw a historic high in the number of instances of terrorist acts or attempted terrorist acts. The Database records a total of 6,660 incidents for that year.[108] For the same year, it records one terrorist incident in the United States and four incidents in Germany.[109] Europol's Terrorism Situation and Trend Report, however, lists a total of 13 attacks in Germany for that year.[110] In contrast, the Federal Trade Commission recorded 20,475 instances of fraud involving credit cards in the United

[106] Ibid.
[107] See note 91 above and accompanying text.
[108] RAND Database of Worldwide Terrorism Incidents, available at http://smapp.rand.org/rwtid.
[109] Ibid.
[110] EUROPOL, "EU Terrorism Situation and Trend Report" (March 2007) at 13, http://www.europol.europa.eu/publications/EU_Terrorism_Situation_and_Trend_Report_TE-SAT/TESAT2007.pdf.

States in that year[111] and over 60,000 instances of identity theft that involved credit card fraud.[112] The German Bundeskriminalamt recorded 8,932 incidents of credit card fraud in Germany for that year.[113] A particularly high incidence of terrorism in the United States is recorded for the year 2001. For that year, the RAND database lists 36 incidents in the United States, one incident in Germany and a total of 1,732 incidents worldwide. Bundeskriminalamt data, however, registers 57,713 incidents of credit card fraud in Germany for that year.[114]

Rarity represents a problem for data mining on a number of levels. For instance, it represents a problem in terms of testing or evaluating the performance of any given data mining exercise. In order to evaluate how well a particular data mining programme performs the task foreseen for it, one has to have a set of test data within which the expected values are known. In this way, one would be able to identify which true positives were correctly assessed as well as identify the number of false positives and false negatives. Additionally, the test data need to be representative of the actual data to which the data mining programme will ultimately be applied. When dealing with phenomena that are rare, there can be less certainty that instances of the phenomenon of which we have records will reflect future instances of the phenomenon which we want to identify. It also means that less information is available on which to build models or profiles.[115] Thus, we cannot have as much confidence in those models and profiles as would be the case with more frequent phenomena. Some applications of data mining may try to compensate for the lack of certain data by extrapolating from existing data. The rarity of terrorist events, however, makes this kind of extrapolation less reliable.[116]

3.4.2. The Problems of False Positives and False Negatives

Any discussion of the effectiveness of data mining will include reference to false positives and false negatives. Broadly speaking, false positives represent results that are shown to meet our criteria but do not in fact represent the results we are truly interested in. Thus, for example, if the task of a data mining programme is to identify terrorists, any individuals who are incorrectly identified as terrorists represent false positives. Conversely, any terrorists within our data set whom the data mining exercise fails to identify as terrorists would represent false negatives. Generally speaking, data mining algorithms can be tweaked to either minimize false positives or minimize false negatives.[117]

---

[111] Federal Trade Commission, "Consumer Fraud and Identity Theft Complaint Data" (February 2008) at 7, http://www.ftc.gov/sentinel/reports/sentinel-annual-reports/sentinel-cy2007.pdf.
[112] Ibid., p. 13
[113] Kriminalistisches Institut, Bundeskriminalamt, "Polizeiliche Kriminalstatistik 2006" (2007) at 186, http://www.bka.de/pks/pks2006/download/pks-jb_2006_bka.pdf.
[114] Kriminalistisches Institut, Bundeskriminalamt, "Polizeiliche Kriminalstatistik 2001" (2002) at 192, http://www.bka.de/pks/pks2001/index2.html. Specific data from the FTC's Consumer Sentinel Reports is not available for that year. However, a total of 220,343 complaints concerning fraud and identity theft were received. Federal Trade Commission, "National and State Trends in Fraud and Identity Theft" (2004) at 4, http://www.ftc.gov/sentinel/reports/sentinel-annual-reports/sentinel-cy2003.pdf. The following year, nearly twice as many complaints were received. Ibid.
[115] See also, e.g., National Research Council, supra note 55, p. 214.
[116] Ibid., p. 214
[117] See, e.g., B. Schneier, *Data Mining for Terrorists*, http://www.schneier.com/blog/archives/2006/03/data_mining_for.html (2006).

However, it may be impossible to ensure that the number of false positives reaches zero unless the test data used to set the algorithm are a very accurate representation of the actual data on which the programme will be used. On the other hand, setting the algorithm to eliminate false positives may introduce more false negatives than are considered acceptable. Thus, from the perspective of false positives and false negatives, counter-terrorism data mining efforts will either point suspicion at innocent individuals or fail to turn up all suspects—perhaps even to a point that calls the benefit of the exercise into question.

Security specialist Bruce Schneier and IT specialist Jeff Jonas together with policy thinker Jim Harper all criticized the notion that data mining could uncover terrorists. All essentially relied on the absence of a well-defined profile due to the rarity of terrorist action and the base rate implications stemming from the large volume of data that would be mined in counter-terrorism programmes. The rarity issue would entail that the number of true positives would be relatively small and would also prevent the reduction of the number of false positives to an acceptable number.[118] Thus, Jonas and Harper argued that a programme with a low false positive rate of only 1% would find 3 million "terrorists" when applied to the entire US population of some 300 million people.[119] Schneier, referring to the Terrorist Information Awareness programme,[120] seems to assume that counter-terrorist data mining programmes would be working with various "event"-related data such as transactional data, e.g., credit card purchases, and communicational activities (e-mail, telephone calls, internet usage, etc.).[121] He draws upon a hypothetical model to illustrate his point:

> We'll be optimistic. We'll assume the system has a 1 in 100 false positive rate (99% accurate), and a 1 in 1,000 false negative rate (99.9% accurate).
>
> Assume one trillion possible indicators to sift through: that's about ten events -- e-mails, phone calls, purchases, web surfings, whatever – per person in the U.S. per day. Also assume that 10 of them are actually terrorists plotting.
>
> This unrealistically-accurate system will generate one billion false alarms for every real terrorist plot it uncovers. Every day of every year, the police will have to investigate 27 million potential plots in order to find the one real terrorist plot per month. Raise that false-positive accuracy to an absurd 99.9999% and you're still chasing 2,750 false alarms per day – but that will inevitably raise your false negatives, and you're going to miss some of those ten real plots.[122]

Schneier, Jonas, and Harper also all refer to the use of data mining in other contexts. Schneier refers to applications to combat credit card fraud, Harper refers to marketing efforts aimed at generating additional sales, and Jonas and Harper refer to customer relationship management—another marketing function. The authors cite these examples to

---

[118] Schneier, supra note 117; J. Jonas and J. Harper, "Effective Counterterrorism and the Limited Role of Predictive Data Mining" (2006) at 7–8; J. Harper, *Data Mining Can't Improve Our Security*, http://www.cato.org/pub_display.php?pub_id=6832 (27 August 2010).

[119] Jonas and Harper, supra note 118, 8. (citing Jeffrey Rosen, *The Naked Crowd* (Random House 2005)). See also R. J. Bolton and D. J. Hand, "Statistical Fraud Detection: A Review" (2002) 17:3, Statistical Science, 235–55 at 236.

[120] See D08.1, pp. 9-10.

[121] Schneier, supra note 117.

[122] Ibid. Schneier's model assumes one trillion events a year among US persons. The 1% false positive rate would mean that there would be ten billion false positives a year or more than 27 million each day.

illustrate how the level of false positives is an acceptable cost in these contexts but not so in the counter-terrorism context. We will discuss this point further in Section 5.

Given the widespread claims of data mining success in private sector applications, it is all the more remarkable that comprehensive studies evaluating data mining performance in terms of false positives and negatives cannot be found. One partial exception is a 1998 article by Abbott, Matkovsky, and Elder which seeks to evaluate the performance of commercial data mining software products that were among the market leaders at that time.[123] Part of their evaluation involved testing the products' performance in fraud detection. Unfortunately, the authors do not provide a particularly thorough account of their methodology in assessing fraud detection performance: For instance, it is not revealed how many individuals or transactions were represented in the test data, nor are we told what the actual number of fraudulent instances were which the ideal software tool would pick out. Clearly, the authors' aims are to provide software procurement agents with an uncomplicated, comparative overview rather than a detailed, empirical assessment. Nonetheless, some interesting inferences can be drawn from their discussion.

The authors compare five data mining products, including two prominent products, Clementine[124] and SAS's Enterprise Miner, which are not unknown in the law enforcement community.[125] For each of the products, the authors applied the product's decision tree and neural network algorithms separately to the test data and recorded the number of false positives and false negatives with each application. The results are interesting in that neither type of algorithm proved to be more effective than the other across the board.[126] For two of the products, however, the use of the decision tree algorithm reduced the number of false positives remarkably over application of the neural networks. The authors explain this occurrence as follows:

> This is probably primarily due to two factors. First, most of the trees allowed one to specify misclassification costs, so nonfraudulent transactions could be explicitly given a higher cost in the data, reducing their number missed. Secondly, pruning options for the trees were somewhat better developed than the stopping rules for the networks, so the hazard of overfit was less. [127]

The authors go on to note, however, that in other contexts they have found "the exact opposite performance."[128]

With respect to false positives in particular, there was significant divergence in performance. The worst performance had just over 20 false positives whereas the best performance had fewer than 5.[129] As for false negatives, the worst performer correctly detected just over 40 fraudulent transactions, whereas the best performer identified a little over 80.[130] Again, we

---

[123] D. W. Abbott, I. P. Matkovsky and J. F. Elder, *An Evaluation of High-end Data Mining Tools for Fraud Detection*, (1998), http://www.abbottanalytics.com/assets/pdf/Abbott-Analytics-Evaluation-High-End-DM-Tools-1998.pdf.
[124] At that time, still a product of the UK company Integral Solutions, Ltd.
[125] See, e.g., McCue, supra note 32, pp. 134 et seq.. Note that the evaluation contained in the article by Abbott et al. may not reflect the performance of current versions or successors of these products, not to mention, customized solutions based on these products.
[126] Abbott, Matkovsky and Elder, supra note 123, pp. 5-6 and Figs. 1 & 2 in particular.
[127] Ibid., p. 5.
[128] Ibid., pp. 5–6.
[129] Ibid., Fig. 1.
[130] Ibid., Fig. 2.

are not told what the correct number of fraudulent transactions was, so we cannot assess how many transactions even the best performer missed. Since the authors do not reveal the total number of transactions within the set of test data or the percentage of fraudulent transactions, it is also difficult to put these numbers into perspective.

Overall, in terms of benefits, the programs had some success in identifying fraud however the authors had defined it within the test data. Generally speaking, the programs returned a higher number of true positives than false positives. However, in some instances the ratio of false positives to true positives was remarkably high and calls into question what true benefit these particular applications would bring to fraud detection efforts. The authors do not reveal how much time was put into developing models and testing. This factor is significant if the use of data mining is to provide some time-saving benefit. We also have to consider the time required to update data mining systems. Frauds may change their techniques and practices over time, in some instances in direct response to deployed detection technologies.

There may be methods for improving the rate of false positives and negatives. David Jensen was one of the academics involved in DARPA's TIA project, and in a 2003 article entitled "Information Awareness: A Prospective Technical Assessment",[131] he together with Rattigan, and Blau outlined an approach which likely reflects the aims of at least one of the programmes within TIA. They argue that their approach would significantly decrease false positives in the counter-terrorism context. There are three elements characterizing this approach which they discuss in the article. The first is reliance on "relational" as opposed to "propositional" data. In short, their approach would not simply apply analysis to a database of entities with certain properties. Rather, relations between the entities would be discovered and defined, permitting subsequent analysis to take these relations into account.[132] In other words, data mining would take place within social networks or networked transactions, etc. The second element is reliance on ranking classifiers as opposed to a simple binary valuation. Binary classification would classify each entity as either "true" or "false", e.g. suspect or not suspect. Ranking classification, on the other hand, would calculate the likelihood that an entity should be labelled as "true" in much the same way as a search engine may rank search results based on its model of determining relevance to a given set of search criteria.[133] Ranking can play a significant role in research and analysis if the rank score or some visual representation of the rank score is available to the user. That visual cue provides the user with an indication of the strength of the connection. The third element of Jensen's approach involves multiple, recursive applications of a data mining algorithm or algorithms to the data.[134] As the authors point out, this sort of "multi-pass" method could even involve different sets of data for each pass or at least less data for initial passes. In this way, the method could potentially provide a certain level of privacy protection by utilizing less personal or sensitive data for the initial pass.[135] Ideally, the use of multiple applications of data mining algorithms would continuously narrow down results.

The authors claim that testing with a system utilizing all three elements demonstrated that the second pass provided a significant improvement in the ratio of true positives to false

[131] D. Jensen, M. Rattigen and H. Blau, "Information Awareness: A Prospective Technical Assessment", in , *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: Association for Computing Machinery, 2003).
[132] See Ibid., pp. 3–4.
[133] See Ibid., p. 4.
[134] Ibid., p. 5.
[135] See Ibid., pp. 5–6.

positives over the initial run. For instance, their figures indicate that at a 10% false positive rate, the initial run had a true positive rate of approximately 50%. At this level, however, the second run shows a true positive rate of nearly 80%.[136] The level of superiority of the second run over the first run diminishes, however, both as the level of false positives increases and decreases. Thus at a 3% false positive rate, the benefit may be only a 10% increase over the first run—and at this level, the true positive rate is less than 50%—and at a false positive rate of 1%, there may be effectively no benefit at all provided by the second run.

It may be tempting to turn back to the critique of Bruce Schneier. Does the test model that Jensen et al. have developed perform better than Schneier's hypothetical data mining programme with a false positive rate of 1% and true positive rate of 99.9% which was the basis of his criticism? The short answer is that Jensen's model would not even remotely match Schneier's ideal programme. Additionally, the 2-cycle model referenced in Jensen's article is obviously a very simple one. It merely serves to indicate that the performance of the model can be improved through subsequent cycles of processing. Jensen et al. appear to be inviting us to conceive of how much better a system would perform if it incorporated five, seven, dozens, or even scores of algorithmic passes. Nonetheless, it seems unlikely that any system will be able to match or surpass the idealized programme that Schneier uses for his thought experiment. Jensen may be likely to counter that Schneier's model assumes binary output (e.g., suspect/ not suspect) as opposed to ranking classification or a suspicion score. Regardless of whether Schneier did have a binary model in mind, it is ultimately irrelevant for his argument. It does not matter how we define a "positive"—whether in terms of binary classification or rank score—false positives are a problem for both. Moreover, we do not know whether Jensen's method performs the same in environments where 1 out of 100 items represent a positive as in environments where 1 out of a million items represents a positive. A multiple-pass system would also inevitably involve the application of more processing power than a single-pass system. That may make that kind of system unfeasible for some environments or applications.

3.4.3. The Problem of Data Quality

Another problem which data mining encounters is that of poor data quality. Issues with the quality of data on which data mining will rely can take a number of forms. Data may be recorded incorrectly due to human error. For instance, someone may inadvertently invert the digits in a number when entering it on a form or a data entry clerk may misspell a name when entering data into a database. Additionally, data may become corrupted through some form of computer error or a compatibility problem. Data may also be missing due to the fact that it was unavailable at the time of entry or the person entering the data simply failed to fill out all the required fields on a form. Methods have been developed to compensate for missing data, but the use of these methods may introduce another source of error.

Problems in the quality of data within private sector databases in the US have been widely reported. For instance, a small survey conducted by the National Association of State Public Interest Research Groups found that out of 197 credit reports from 154 adults in 30 different states, 79% contained some sort of error.[137] Over half of the reports contained inaccuracies

[136] Ibid., pp. 7, Fig. 3.
[137] National Association of State PIRGs, "Mistakes Do Happen: A Look at Errors in Consumer Credit Reports" (June 2004) at 13 & 16.

with regard to personal information, such as misspelling of names, incorrect birth dates, outdated addresses listed as current addresses, or addresses listed where the individual had never lived.[138] According to the report, some of these kinds of errors result from "mis-merges", where information about one individual is mistakenly added to a file pertaining to a different individual.[139] Additionally, the report alleges that errors may also be traced to the failure on the part of credit reporting agencies to verify the identity of individuals when incorporating information from public records. Thus, a bankruptcy filing submitted in the name of one John Smith may end up in the credit report of another John Smith.[140] Once these errors are generated in a source database they can be propagated to additional databases including those maintained by law enforcement and other government agencies. As was revealed in D08.1, several data mining systems in the US that are used in the counter-terrorism context rely on data obtained from commercial sources. One such source might be ChoicePoint (now LEXIS-NEXIS) databases, which the FBI has certainly had access to.[141] Another data aggregation business, Axciom, may have been tapped as a potential partner for TIA or individual projects within TIA.[142] These private data aggregation companies in turn obtain their data from a variety of public and private databases, likely including data from credit reporting agencies. The spread of erroneous data among different databases also produces an issue for the rectification of that data; correcting the data in one database will often not result in the correction of the data in other databases. With the increasing proliferation of automated aggregation services,[143] not only can erroneous data be further propagated, but these services may add new errors as well.[144]

3.5. Summary

Data quality and false positives and negatives raise issues for data mining generally. More specifically, the relative rarity of terrorist events raises problems for modelling terrorist activity. One message that can be derived from both the Abbott and Jensen articles is that false positives can be reduced. Nonetheless, they will always be a problem.[145] Abbott and other authors stress the importance of testing, modelling, and updating in order to develop data mining regimes that are reliable and effective. Although in a few instances independent bodies have been invited to review the development of IT-based security

[138] Ibid., p. 12. The problem of misspelled names may be particularly accute in the counter-terrorism context where the names of suspects may be unfamiliar or uncommon from the perspective of many Westerners or derived from languages that do not rely on the Roman alphabet.
[139] Ibid., p. 6.
[140] Ibid., p. 7.
[141] See, e.g., Office of the Inspector General, U.S. Department of Justice, "A Review of the FBI's Handling of the Brandon Mayfield Case" (2006) at 35. Note that Mary de Rosa also suggests that these commercial databases would provide key data for effective identity reseolution. DeRosa, supra note 44, p. 10.
[142] See the Electronic Privacy Information Center's TIA webpage at http://epic.org/privacy/profiling/tia/ and the corresponding document containing e-mail correspondance in response to a Freedom of Information Act request at http://www.epic.org/privacy/profiling/tia/darpaacxiom.pdf.
[143] Examples of such services that aggregate content from public records and other open sources such as social networking sites include Yasni, 123people, zoominfo, Intelius (used by yahoo.com search), USSearch, peoplefinders, peoplesmart, whitepages.com, and spokeo.
[144] See also K. A. Taipale, "Designing Technical Systems to Support Policy: Enterprise Architecture, Policy Appliances, and Civil Liberties", in R. Popp and J. Yen (eds.), *Emergent Information Technologies and Enabling Policies for Counter-Terrorism* (IEEE Press, 2006), p. 454 on the problem of "repurposing" data that was collected under low standards of accuracy.
[145] See Schneier, supra note 117.

measures aimed at least in part to counter terrorism and have provided some documentation of their oversight activities, in the vast majority of instances there is no evidence that such programmes have received any testing of consequence.[146] Detailed figures such as false positive and false negative rates have not been released to the public. We should also be wary of the assumption that a particular system will function more or less equivalently in the operational environment as it did in the test environment.

In addition to a lack of comprehensive figures pertaining to the application of data mining in any setting, we noted in Section 3.4 above that there is little apart from isolated anecdotes to speak for the benefits of the application of data mining in the counter-terrorism context. We now turn to an assessment of the costs connected with data mining.

4. Costs

There a number of different classes of cost associated with data mining. In the Introduction above, we mentioned a number of—to use a term from economics—"direct costs" associated with developing and implementing data mining systems in terms of financial costs, human resources, etc. Opportunity costs of a negative kind may also arise when resources dedicated to data mining and following up on the analysis resulting from data mining prevents authorities from pursuing more fruitful counter-terrorism efforts. In addition, there are costs that flow from the occurrence of false positives. These include costs in terms of following up on false positives as well as costs in terms of the infringement of the rights of those implicated as false positives. Furthermore, rights infringements may also occur even in the absence of false positives, depending on the privacy and data protection rights that may be recognized in the jurisdiction where the data mining takes place or where the subjects of data mining reside or are citizens. We will divide these various costs into two groups for purposes of the following discussion: those costs which accrue to government agencies and those which accrue to the subjects of data mining exercises—i.e. the impact in terms of human rights violations.

4.1. Costs for Government Agencies

As noted above, one way in which data mining exercises may create costs for the government agencies that rely on them is through the poor allocation of resources. The German *Rasterfahndung* provides one example where the use of data mining had such a result. As part of that exercise, numerous police officers and potentially other public servants throughout the country as well were reassigned from their usual duties to collect, compile, and analyse data for this profile-based search. It is unknown to what extent the initiative interfered with the usual operations of those non-law enforcement agencies that were involved. At least with respect to North Rhine-Westphalia, however, crime statistics suggest that time invested in carrying out the *Rasterfahndung* had a marked negative impact on law enforcement in other areas. There, 600 officers were assigned to work on the initiative over the space of several months.[147] During this time, there was an increase in the

[146] This finding echoes remarks made by RAND president James A. Thomson in 2007: "...DHS implements most of its programs with little or no evaluation of their performance. When performance metrics have been implemented, they have often measured inputs and outputs only — not effectiveness." J. A. Thomson, "DHS AWOL?: Tough Questions About Homeland Security Have Gone Missing" (Spring 2007) 31, RAND Review, 30.
[147] Berndt and Schiffermüller, supra note 54.

number of common crimes such as muggings, break-ins, etc.[148] Notably, in this instance, there was no specific terrorist threat that prompted the exercise. It may have been possible to automate aspects of the *Rasterfahndung* to avoid tying up so many law enforcement personnel; however, even assuming this possibility existed, there would have been financial costs in terms of designing the system to perform these functions that would take the place of the time/attention costs of police investigators. Moreover, as mentioned above, no terrorist suspects were identified as a result of the initiative. There is also no indication that a looming terrorist threat manifested itself due to the failure of the *Rasterfahndung*, and whether the exercise had any deterrent effect can only be left to speculation. Overall, the programme resulted in a net loss for public safety.

There are of course also the financial costs of developing and operating data mining programmes. It is difficult to get comprehensive figures for the cost of data mining programmes being applied for counter-terrorism purposes; however, numbers are available for components of the TIA project. For the Evidence Extraction and Link Discovery (EELD) project, for example, budget statements indicate that 17.344 million USD were spent on the project in 2001,[149] 12.309 million USD were spent in 2002.[150] In 2003, the budget projections for the years 2003 – 2005 totalled 32.332 million USD.[151] Thus, at least 29 million USD were spent on research and development for this component before funding was officially withdrawn by Congress and a total exceeding 41 million USD had been budgeted over a period of 4 years. The TIDES and EARS projects survived Congressional defunding. In 2002, 27.831 million USD was spent on those projects[152] and 34.141 million USD was spent in 2003.[153] At the beginning of 2004, an additional 46.332 million USD was allocated to the projects,[154] suggesting that ultimately, over 100 million USD was disbursed on the projects.

The ADVISE programme reportedly had an overall proposed budget of approximately 42 million USD for the years 2003-2007.[155] As noted above, this programme was abandoned due to the fact that development expenditures were no longer considered justified in light of the availability of comparable and less expensive commercial, off-the-shelf products. Using off-the-shelf solutions will certainly avoid the kinds of development costs associated with programmes like EELD and TIDES and EARS. However, it is likely to be a rare case where existing commercial products can perform the kinds of specialized tasks required in counter-terrorism. Additionally, there may still be service costs associated with the integration of the software into the agency's operational environment as well as custom add-ons to address legal compliance issues that commercial users may not face.

---

[148] Berndt and Schiffermüller, supra note 54 (statement of Wilfried Albishausen, Bund der Kriminalbeamter (Union of Criminal Justice Officers)).
[149] U.S. Department of Defense, "Fiscal Year (FY) 2003 Budget Estimates: RESEARCH, DEVELOPMENT, TEST AND EVALUATION, DEFENSE-WIDE, Volume 1 - Defense Advanced Research Projects Agency", (2002) at 90.
[150] U.S. Department of Defense, "Fiscal Year (FY) 2004/FY 2005 Biennial Budget Estimates: RESEARCH, DEVELOPMENT, TEST AND EVALUATION, DEFENSE-WIDE, Volume 1 - Defense Advanced Research Projects Agency", (2003) at 70.
[151] Ibid.
[152] Ibid., p. 33.
[153] U.S. Department of Defense, "Department of Defense Fiscal Year (FY) 2005 Budget Estimates: RESEARCH, DEVELOPMENT, TEST AND EVALUATION, DEFENSE-WIDE, Volume 1 - Defense Advanced Research Projects Agency", (2004) at 61.
[154] Ibid.
[155] Office of the Inspector General, Department of Homeland Security, "ADVISE Could Support Intelligence Analysis More Effectively", OIG-07-56 (2007) at 7.

Under certain circumstances, data mining can actually exacerbate the problem of information overload and lead to analytic impairment. There are also data quality issues that do not stem from data entry errors, corrupted data, or omissions but rather from poor definition of the purpose of programmes and accompanying data as well as poor coordination among agencies that build and make use of the information environment. As information becomes more widely distributed, whether as the result of automated services or deliberate sharing on either an ad hoc or systematic basis, there is the danger not only that erroneous information proliferates but also that information that was arguably correct to begin with is later interpreted in an inappropriate manner. This problem can arise, for example, where systems that were originally "stovepiped" later become accessible to other organizations or agency units, but also where analysts who are authorized to add information do not know or contemplate how that information may be used downstream. In both situations, individuals who come across the data often will not know where that data comes from or why it is in the system to begin with. Moreover, the mere fact that someone has received attention from intelligence or law enforcement may be taken as incriminating.

A case in point is that of Maher Arar. Arar is a dual Canadian-Syrian citizen, who was identified as a person of interest by the Royal Canadian Mounted Police (RCMP) due to the fact that he had known associations with another individual believed to be affiliated with al Qaeda. The RCMP requested of Canadian and US Customs that both Arar and his wife be placed on "lookout" lists and that they were "'Islamic Extremist individuals suspected of being linked to the Al Qaeda terrorist movement.'"[156] These notifications likely did not give the impression that Canadian authorities did not have any reason to arrest or charge Arar or his wife with any terrorism-related offenses. Other information concerning Arar that Canadian authorities provided to their US counterparts prior to Arar's stop at JFK Airport likely compounded the problem by including statements that referred to Arar as a "suspect", "principal subject", "target or important figure" and that suggested that Arar had refused to be interviewed by the RCMP.[157] As the Canadian Commission of Inquiry that examined the Arar matter noted, in many instances, "no explicit caveats were attached to the information sent to the Americans."[158] Although Canadian officials later made it clear to FBI agents that they had yet to establish definitive ties to al Qaeda,[159] at that stage, the FBI seemed convinced that Arar was a terrorist.[160] As can also be seen in, for instance, the case of

[156] Rendition to Torture: The case of Maher Arar: Joint Hearing Before the Subcommittee on International Organizations, Human Rights, and Oversight of the Committee on Foreign Affairs and the Subcommittee on the Constitution, Civil Rights, and Civil Liberties of the House Committee on the Judiciary, 110th Cong., 1st Sess. (2007) (Testimony of Maher Arar), p. 1. Rendition to Torture: The case of Maher Arar: Joint Hearing Before the Subcommittee on International Organizations, Human Rights, and Oversight of the Committee on Foreign Affairs and the Subcommittee on the Constitution, Civil Rights, and Civil Liberties of the House Committee on the Judiciary, 110th Cong., 1st Sess. (2007), Transcript, p. 44 (Prepared Statement of Kent Roach, Esq.) (citing Commission of Inquiry into the Actions of Canadian Officials in Relation to Maher Arar Factual Background Vol 1 (2006) at 62).
[157] Commission of Inquiry into the Actions of Canadian Officials in Relation to Maher Arar, "Report of the Events Relating to Maher Arar: Factual Background, Vol I" (September 2006) at 113, http://epe.lac-bac.gc.ca/100/206/301/pco-bcp/commissions/maher_arar/07-09-13/www.ararcommission.ca/eng/Vol_I_English.pdf.
[158] Ibid., p. 114.
[159] Ibid., p. 160.
[160] It is unclear whether US intelligence had additional information on Arar which the Canadians did not. Although the Canadians passed on their work and findings concerning Arar, the Canadians never received any information concerning Arar from their US counterparts that convinced them of Arar's involvement in al Qaeda. See Ibid., pp. 114–16. The only suggestion that the FBI had additional information was a conversation reported by then Canadian Minister of Foreign Affairs Bill Graham in which he was told by then Secretary of State Colin Powell that the US knew of phone numbers and a telephone call "that justified deporting Mr. Arar". Ibid., p. 298. Statements made by then US

Brandon Mayfield[161] as well as other well-known instances, compounding circumstantial evidence can act to create a spiral of increasing suspicion, which over time, becomes increasingly difficult to overcome.

The Mayfield case provides a notable example of how the use of technology can mislead human users. As the OIG noted in its report on the FBI's handling of the case, the automated fingerprint matching system that the FBI employed was designed precisely to present the examiner with numerous candidates that most closely resemble the print of interest. Thus, in an ideal case, the system would "find not only the source of the print (if it is in the database), but also the closest possible non-matches."[162] Thus, if there are prints which are remarkably close to one another, the system essentially returns those fingerprints that are most likely to confuse the examiner.[163] The OIG thus concluded that "[t]he likelihood of encountering a misleadingly close non-match through an IAFIS search is therefore far greater than in a comparison of a latent print with the known prints of a suspect whose connection to a case was developed through an investigation."[164] The OIG interviewed one individual who had served for 14 years on the IAI Certification Board, which was responsible for investigating complaints of erroneous identifications by IAI-certified examiners. This individual indicated that he had encountered 25 to 30 erroneous fingerprint identifications and all but one of those cases, involved the use of automated matching programmes.[165] The OIG also described how a process of "circular reasoning" set in which, once Mayfield's fingerprint was presented as a possible match, examiners looked for features in the latent print that could arguably be matched to Mayfield's print rather than remaining strictly on a path of analysis in the other direction – i.e. proceeding from the latent print to the possible matches.[166] In other words, the OIG suspected that examiners allowed the suggested match to bias their analysis and interpretation of the latent print.

The kinds of costs relating to analytic failures that are described here in many instances lead in turn to costs for individuals in the form of human rights infringements.

4.2. Costs for Individuals

In terms of human impact, there are at least two categories of risks which the use of data mining entails: one is the risk that security forces will take action against individuals implicated through false positives; the other is the inherent risk to privacy stemming from the collection and handling of personal data in digital form. A third form of risk which may arise from some forms of data mining is the risk that innocent parties are drawn into an investigation due to links to terrorist suspects. Related to the first risk is the danger that analyses delivered through the application of information technologies will prove too convincing or appealing for human analysts to view critically or otherwise "shortcircuits" the analytic process.[167] Additionally, human bias may play a role in this process, whether

---

Ambassador to Canada Paul Cellucci suggested that the reasons for removing Arar to Syria were based on the information that had been provided by the Canadians. Ibid., p. 232.

[161] See, e.g., S. T. Wax and C. J. Schatz, "A Multitude of Errors: The Brandon Mayfield Case" (2004) September/October, The Champion, 6; Office of the Inspector General, U.S. Department of Justice, "A Review of the FBI's Handling of the Brandon Mayfield Case" (2006).

[162] U.S. Department of Justice, supra note 161, 137.

[163] Ibid., p. 137.

[164] Ibid., p. 137.

[165] Ibid., p. 137.

[166] Ibid., pp. 138–50.

[167] See D03.1, p. 4 concerning the seductive danger of the 'pretty picture'.

consciously or unconsciously, where technological solutions appear to confirm what one "knew all along".[168] Results presented by technology can also bias the reasoning process such that the analyst seeks out details to confirm the result rather than proceeding from facts to a conclusion.[169] Bias may also lead in the other direction—where an analyst ignores a true positive because it does not conform to the analyst's idea of a true suspect. A report within the Future of Identity in the Information Society Project noted an instance in which the activities of a particular Goldman Sachs employee had triggered fraud alerts on several occasions, but the individual was not initially investigated since she did not belong to a "socio-demographic group" that was typically involved in money laundering.[170]

Data mining is often used as a means of prioritizing attention.[171] In this way, it focuses the scrutiny of analysts and investigators on particular individuals implicated in query results. Where this scrutiny concerns a false positive, the action, at the least, constitutes an undesirable infringement of that individual's privacy. In some contexts, it may also entail confrontation from security personnel, such as in the cross border context. Data mining may also be used to aggregate data from various sources. When this occurs, the aggregate data can provide a much richer picture of an individual's personal life, family, financial affairs, interests, and activities than that data would reveal in distributed form even if it were publicly accessible. Additionally, such aggregating functions can pose a greater danger for breaches or leaks of data as discussed below.

Data handling risks include the risk of the exposure and transfer of data through both insiders—i.e. those who have legitimate access to the data—as well as outsiders—often popularly referred to as "hackers". The US Secret Service together with Carnegie Mellon's Software Engineering Institute published a report on IT-related insider sabotage which references a number of instances in which insiders—often disgruntled employees or former employees—sabotaged private and public computer systems and misappropriated critical data or software.[172] The recent Wikileaks compromise of information held on US Army Intelligence systems demonstrates how easily large amounts of sensitive data can be transferred beyond its intended setting when proper safeguards are not in place.[173] Even civilian intelligence and law enforcement agencies have seen a number of high-profile insider breaches—such as Robert Hanssen (FBI)[174] and Aldrich Ames (CIA).[175] The problem,

---

[168] This is not to suggest that "gut feelings" or experiential knowledge should play no role in assessing data mining results.
[169] An example of this "reverse reasoning" process can be found in the Brandon Mayfield case.
[170] Future of Identiy in the Information Society (FIDIS), D7.2, Descriptive analysis and inventory of profiling practices, M. Hildebrandt & J. Backhouse (eds.), p. 58.
[171] M. DeRosa, *Data Mining and Data Analysis for Counterterrorism* (26 June 2009), p. 6.
[172] M. Keeney, E. Kowalski, D. Cappelli, A. Moore, T. Shimeall and S. Rogers, "Insider Threat Study: Computer System Sabotage in Critical Infrastructure Sectors" (May 2005).. See p. 3 in particular.
[173] See, e.g., K. Poulsen and K. Zetter, "U.S. Intelligence Analyst Arrested in Wikileaks Video Probe", (6 June 2010), Wired, Threat Level, available at http://www.wired.com/threatlevel/2010/06/leak/.
[174] See, e.g., U.S. Department of Justice, *A Review of the FBI's Performance in Deterring, Detecting, and Investigating the Espionage Activities of Robert Philip Hanssen* (Washington, D.C.: U.S. Government Printing Office, August 14, 2003). The Hanssen case is particularly interesting in that the FBI, cognizant that there was a leak somewhere, became convinced that CIA agent Brian Kelley was the culprit despite the fact that a number of alarms had been raised with respect to Hanssen. Kelley was placed under FBI surveillance and at least one attempt was made to try to ferret him out with deceptive tactics. Ultimately, he was suspended from work for an extended period until Hanssen was finally caught. See "To Catch A Spy: Probe To Unmask Hanssen Almost Ruined Kelley" (30 Jan. 2003), 60 Minutes, available at http://www.cbsnews.com/stories/2003/01/30/60minutes/main538650.shtml?tag=mncol;lst;4.
[175] See, e.g., U.S. Senate, Select Committee on Intelligence, *An Assessment of the Aldrich H. Ames Espionage Case and Its Implications for U.S. Intelligence* [84-046] (Washington, D.C.: U. S.

of course, is not unique to the colourful realm of international espionage.[176] And government agents who have access to sensitive data could potentially pocket much higher sums through a single sale than Hanssen supposedly gained through his many years of selling secrets. Reportedly, black market websites already exist that offer a variety of personal data[177] and there are even "legitimate" markets where foreign governments are willing to pay for data.[178]

In addition to the intentional appropriation of data, data can also be exposed unintentionally through the actions of employees or contractors. Particularly in the UK, there have been a number of high-profile incidents in which public sector data has been lost or stolen: For instance, in 2007, a hard drive containing data on 5,000 justice ministry personnel was lost by a private firm.[179] In that same year, two discs containing data on all claimants of UK child benefits that Revenue and Customs was attempting to send to the National Audit Office were lost in the mail.[180] The private sector has also not been immune to data loss and theft. In both the US and the UK, there have been numerous high-profile data breaches involving commercial firms. One of the most prominent was the TJX incident in the US, which concerned the appropriation of credit and debit card details by exploiting vulnerabilities in

---

Government Printing Office, 1994).

[176] See, e.g., U.S. v. Nichols, 820 F.2d 508 (1st Cir.1987) (customs agent caught selling information from the Treasury Enforcement Communications system); M.D. Simon, "Cops suspended for Obama check" (30 July 2009), Atlanta Journal-Constitution, http://www.ajc.com/news/dekalb/cops-suspended-for-obama-103821.html (local police officers run unwarranted background check on U.S. President Obama); see also B. Flora, "What Do the Cops Have on Me?" (4 December 2007), Slate, http://www.slate.com/id/2179180/ (quoting an attorney who claimed that it was commonplace within one local police agency for officers to "run [background] checks for friends and family, and to run prank names to alleviate boredom.").

[177] One example was the Shadowcrew website. See Indictment, US. v. Mantovani, (D. N.J.) pp. 5-6, available at http://www.usdoj.gov/usao/nj/press/files/pdffiles/firewallindct1028.pdf#search=%22firewallindct1028.pdf%22.

[178] Recently, one individual, who saved various client data from the Liechtenstein high-wealth private bank LGT onto DVDs, reportedly received 5 million Euro from the German government alone. News accounts suggest that he was also able to sell the data to the governments of 12 additional countries. See, e.g., C. Teevs, "Staatsfeind Nummer eins rächt sich" (10 August 2010), *Spiegel Online*, available at http://www.spiegel.de/wirtschaft/soziales/0,1518,711069,00.html; C. Budras, "Ein folgenschwerer Datenklau", FAZ.Net, 07 September 2008, http://www.faz.net/s/Rub53B6D88BDF4A49D6BF5E114728883FE3/Doc~E78D9314AE0E249BDBFFF72CAC0699936~ATpl~Ecommon~Scontent.html; "BND zahlte fünf Millionen für geheime Steuerdaten (16 February 2008), *Spiegel Online*, available at http://www.spiegel.de/wirtschaft/0,1518,535687,00.html; "Tax Whistleblower Sold Data to the US" (25 February 2008), *Spiegel Online International*, available at http://www.spiegel.de/international/business/0,1518,537640,00.html.

[179] "Data on 5,000 justice staff lost", BBC News, 07 September 2008, http://news.bbc.co.uk/2/hi/uk_news/7602402.stm.

[180] "UK's families put on fraud alert", BBC News, 20 November 2007, http://news.bbc.co.uk/2/hi/uk_news/politics/7103566.stm. For additional incidents, see M. Broersma, "NHS top culprit as UK data breaches exceed 1,000" (1 June 2010), http://www.zdnet.co.uk/news/compliance/2010/06/01/nhs-top-culprit-as-uk-data-breaches-exceed-1000-40089098/ ("The NHS has reported 305 breaches since November 2007, according to the Information Commissioner's Office's (ICO) figures. Of those, 116 were due to stolen data or hardware, 87 were due to lost data or hardware and 43 cases were disclosures due to error…."); OUT-LAW.COM, "Trade body loses laptop full of driving conviction data", The Register, 21 August 2009, http://www.theregister.co.uk/2009/08/21/trade_body_data_policy/; R. Winnett and J. Swaine, "Data on 130,000 criminals lost", The Telegraph, 22 August 2008, http://www.telegraph.co.uk/news/newstopics/politics/2601056/Data-on-130000-criminals-lost.html.

the wireless networks of retail stores that were operating as part of the payment approval process.[181]

Including friends, family, and associates of a suspect in the investigation of that suspect is surely a natural, logical, and well-established aspect of investigative methodology. However, there is always the danger that investigators will place these other individuals into the category of suspects without a sufficient basis for doing so. A frequent aim of data mining programmes used in the counter-terrorism context is to perform the task of link analysis—revealing not only these associations but providing some indication or evaluation of the nature and strength of these ties. It must be kept in mind that data quality issues can inevitably have an impact in this context: Incorrect or erroneous data may create false connections; but also missing data or limits on the scope of data included in analysis—whether intentional or unintentional—may provide a much different picture of the strength and significance of relationships than would be the case if more or other types of data were considered. For instance, one terrorist suspect, John, may have frequent communications with another individual, Bob. However, John's far less frequent communications with a third individual, Mary, may be far more interesting if it were also known that these communications closely correlated with known terrorist-related events.[182]

The danger of guilt by association has been implicated in several anti-terrorism cases. In the Brandon Mayfield case, the fact that Mayfield had represented a convicted terrorist in a legal matter, attended the same mosque that other convicted terrorists had attended, and had advertised his law office with a web service that had some purported ties to terrorist figures or terrorist organizations were all evidently considered factors probative enough to be included on the affidavit submitted in support of the warrant for his arrest. In the case of Maher Arar, his acquaintance with a suspected terrorist was apparently the sole basis for his designation as a terrorist suspect. Another individual in California was placed under GPS surveillance for a period of 3-6 months. This individual's association with someone to whom a suspicious web posting was attributed appeared to play a significant role in his being placed under surveillance.[183]

Data quality issues of various kinds can also result in consequences for innocent individuals. One such problem is poor entity resolution. Entity resolution refers to the process of establishing, for instance, that "Osama bin Laden" and "Usama Bin Laden" refer to the same person, that when Osama travels with a passport with the name "Robert Johnson" he is nonetheless in fact Osama bin Laden, and that Osama bin Laden, age 19 of Dayton, Ohio is not the same person as Osama bin Laden, age 42 of the United Arab Emirates and that neither of them is the Osama bin Laden who is likely to be on a watchlist.

The problems of poor entity resolution have been particularly prominent in watchlist matching as applied in the field of aviation security. Numerous failures in the system have been widely reported. Mikey Hicks is one of many individuals who have the misfortune of sharing a name with someone who has been included on flight watchlists. As a result, Hicks has always undergone additional screening at the airport, beginning at the age of 2 when he

[181] See, e.g., Indictment, U.S. v. Gonzalez, No. 08-CR-10223, (D. Mass. 2008).
[182] See, e.g., McCue, supra note 32, pp. 9–10.
[183] K. Zetter, "Caught Spying on Student, FBI Demands GPS Tracker Back", (7 October 2010), Wired, http://www.wired.com/threatlevel/2010/10/fbi-tracking-device; See also "Student's Car Secretly Tracked For More Than 3 Months", Huffington Post, 12 October 2010, http://www.huffingtonpost.com/2010/10/12/yasir-afifi-car-tracked_n_758668.html.

was subjected to frisking by airport security personnel. He was reported as being 8 years old in January 2010 and was still being routinely stopped at security at that time.

The US television news magazine 60 Minutes claimed to have obtained a copy of the "No Fly" List that airport security relies upon in screening passengers. The list was found to contain a number of common names, including "Robert Johnson", "Gary Smith", and "John Williams".[184] The news programme interviewed 12 individuals named Robert Johnson who confirmed that they have all had problems boarding flights.[185] The ACLU has compiled a list of some of the names on the list and individuals affected by it.[186] They include commercial airline pilots, former and current members of Congress, former government officials, and US military personnel. Former South African leader, Nelson Mandela was also on the list until it was removed through an act of Congress.[187]

Sister Glenn Anne McPhee, secretary for education of the US Conference of Catholic Bishops was also inconvenienced at the airport in the years 2003-2004. She learned that the surname "McPhee" was on the list since a certain Afghani had used the name as an alias.[188] She reported routinely missing flights due to the extra security measures and was delayed "up to five hours" at a time.[189] Only after the head of the US Conference of Catholic Bishops wrote a letter to Karl Rove, then Deputy Chief of Staff to President George W. Bush, in 2004 were the issues resolved. A 2005 audit by the Department of Justice Inspector General also determined that names that should have been included on the lists had not been.[190] The Secure Flight programme was introduced in part in the hope of eliminating the kind of mismatches that occurred in the previous system that focused primarily on names.
The current US passenger screening system, Secure Flight, has added gender and data of birth as additional identifiers in the hope of preventing the kinds of false identifications that have occurred in the past.[191] Time will tell how successful the new system will prove in avoiding these kinds of problems.

Data quality errors can also result through poor collection practices where the collection itself is misguided. This can lead to systemic problems where persons and organizations mentioned in intelligence reports are more or less automatically incorporated into databases that carry actionable consequences—such as watchlists. A recent report from the Inspector General to the U.S. Department of Justice describes how individuals related to FBI terrorism investigations may be entered in the FBI's Violent Gang and Terrorist Offender File (VGTOF) and also how policies governing entries to this database changed repeatedly between 2002 and 2005.[192] These entries are then made available to local law enforcement so that local police will be alerted during routine traffic stops or other encounters. Similarly, the FBI's consolidated terrorist watchlist, established in 2004, automatically feeds data to

---

[184] CBS News, "Unlikely Terrorists On No Fly List", (10 June 2007), 60 Minutes.
[185] Ibid.
[186] American Civil Liberties Union, *Unlikely Suspects*, http://www.aclu.org/technology-and-liberty/unlikely-suspects.
[187] Ibid.
[188] R. Singel, "Nun Terrorized by Terror Watch", (26 September 2005), Wired, http://www.wired.com/politics/security/news/2005/09/68973.
[189] Ibid.
[190] M. Sherman, "Review of Terrorism Database Finds Flaws", Washington Post, 14 June 2005, http://www.washingtonpost.com/wp-dyn/content/article/2005/06/13/A.
[191] A system of redress has also been introduced as an additional means to resolve false identifications, for instance, where name, date of birth, and gender might be the same for two different individuals.
[192] Office of the Inspector General, U.S. Department of Justice, "A Review of the FBI's Investigations of Certain Domestic Advocacy Groups" (September 2010) at 25.

numerous other databases, including the VGTOF. Entries in the consolidated watchlist affect screening procedures for commercial flights in addition to other contexts.[193]

An example of poor collection practices is illustrated by the TALON programme. There, the purpose of the programme was to register and distribute information pertaining to threats to military installations. However, what constituted a "threat" came to be interpreted too loosely. Essentially any activity which sought to protest the existence or actions of the military or US involvement in war was deemed to be a threat deserving a report in the database. Additionally, the system for inputting reports may have been too lenient for widespread information-sharing since reports were also accepted from the general public.

A parallel development to the TALON system can be seen in the FBI's targeting of certain civil society groups involved in protest activity. Some of the activities of the FBI were picked up by the media and eventually the Office of the Inspector General of the Justice Department launched an investigation. One of the primary factors that contributed to the controversial practice was the rather broad definition of domestic terrorism that the FBI had adopted. For this reason, the OIG determined that FBI agents had not violated existing policies in classifying certain protest groups or particular individuals as "terrorist"; however, it questioned whether the classification was appropriate since it resulted in crimes being classified as terrorism, which one ordinarily would not associate with terrorism—including such things as vandalism and trespassing. The OIG noted that "[t]he domestic terrorism classification had impact beyond any stigma resulting from the public release of [FBI] documents.... For example, persons who are subjects of domestic terrorism investigations are normally placed on watchlists, and their travels and interactions with law enforcement may be tracked."[194] In the course of that investigation, the OIG determined that several individuals implicated in the cases examined were placed on watchlists.[195] The OIG also noted that placement on a watchlist can have consequences not only for the listed individual, but also for their associates.[196]

With increased information sharing, poor coordination among different agencies also afflicted watchlisting efforts. A 2008 audit of the terrorist watchlisting process that was conducted by the Inspector General to the US Department of Justice found that the National Counterterrorism Center (NCTC), which reviews recommendations for terrorist watchlists, had incorporated information about individuals featured in intelligence reports from the FBI as well as information shared by other federal agencies without the knowledge of either the FBI or any of the other agencies concerned and in spite of the fact that none of the information was intended as a recommendation for inclusion of individuals on a watchlist.[197] Additionally, the audit also discovered that in some instances, FBI field offices were reporting watchlist recommendations directly to the NCTC rather than to FBI headquarters which meant that the quality control function which headquarters could provide was being bypassed.[198]

---

[193] Ibid., pp. 25–26.
[194] Ibid., p. 188.
[195] Ibid., p. 27.
[196] Ibid., p. 27.
[197] Office of the Inspector General, U.S. Department of Justice, "Audit of the U.S. Department of Justice Terrorist Watchlist Nomination Processes", Audit Report 08-16 (March 2008) at 3–4, http://www.justice.gov/oig/reports/plus/a0816/final.pdf.
[198] Ibid., p. 3.

5. Weighing Costs and Benefits

In assessing the benefits of the application of data mining programmes in the counter-terrorism context against both the unavoidable and potential costs, it bears repeating the main findings that were mentioned in Section 3.4 above. First, we have found little to no publicly available evidence demonstrating the efficacy of data mining programmes in the counter-terrorism environment.[199] Second, the best documented example, the German terrorist *Rasterfahndung*, can safely be labelled a clear failure from the information that has been disclosed. However, the *Rasterfahndung* likely represents a particularly poor methodology, and it is possible that better methods are applied in other data mining efforts currently in use.

It is also important to recognize the differences between the use of data mining in the private sector and its potential use in the counter-terrorism context. Some have wanted to point to the use of data mining in other contexts as an indication of its appropriateness for the counter-terrorism context. Yet, as several authors have pointed out, in the commercial setting, the consequences of false positives for human subjects are often relatively trivial.[200] For instance, in the context of credit card monitoring, it would mean an unnecessary phone call from the credit card company to verify the authenticity of a particular transaction. In the context of a marketing campaign, it might mean the receipt of unwanted junk mail or "spam". In the context of counter-terrorism, however, it could mean being barred from international travel or being placed under government surveillance or even arrested.

There are also significant differences with respect to the benefits of data mining in the private sector as opposed to the counter-terrorism context. A targeted marketing campaign may prove "successful" in terms of raising sales enough to more than offset the costs of the campaign despite the fact that a significant portion of those individuals targeted do not react to the campaign.[201] In the anti-fraud context, data mining is used to prioritize attention. It is accepted from the outset that 1) not all positives can be investigated, and 2) not all investigations will lead to a determination of fraud.[202] Due to the large amount of money at stake, however, devoting resources to anti-fraud investigations makes economic sense.[203]

The perceived threat, however, can play a significant role in the assessment of the benefits of data mining. Some will argue that the potential damage from a terrorist attack could

---

[199] Roger Clarke made a similar observation concerning biometric systems in 2002: "What is quite remarkable is the paucity of published, independent assessments of technologies, products and applications. The few commercial assessors demand very high fees for their publications, and governmental testing laboratories and government-funded university laboratories appear to publish very little. Perhaps they are be [sic] embarrassed at the low quality of the things they test; or fearful that, should the poor quality of the technologies in use become publicly evident, they would undermine the credibility of the industry on which their very existence depends; or even concerned about being accused of being unpatriotic if they are perceived to doubt the wisdom of installing technology in order to fight the 'war on terrorism'." R. Clarke, *Biometrics' Inadequacies and Threats, and the Need for Regulation*, http://www.rogerclarke.com/DV/BiomThreats.html (26 June 2009), paras. 2.3.

[200] J. Harper, *Data Mining Can't Improve Our Security*, http://www.cato.org/pub_display.php?pub_id=6832 (27 August 2010); Schneier, supra note 117.

[201] According to Jim Harper, "the direct marketing industry achieves response rates ranging from 5.78 percent for telephone solicitation to 0.04 percent for direct response television." Ibid.

[202] See R. J. Bolton and D. J. Hand, "Statistical Fraud Detection: A Review" (2002) 17:3, Statistical Science, 235–55 at 236–7.

[203] See Ibid., p. 236.

reach catastrophic levels, far beyond that of 9/11—for instance, if terrorists detonated a "dirty bomb". From this perspective, so long as data mining can add even the slightest value toward the prevention of such an attack, its deployment is justified. The real violation of human rights as well as the potential harms that might result may, in comparison, seem minor or even trivial. This argument, however, does not reflect an accurate view of the risk of terrorist harm. Estimates of the likelihood of such a serious attack place it as very remote although not impossible.[204] This argument asks us to trade actual harm for the unsubstantiated possibility of avoiding an unspecified and hypothetical threat. Moreover, such an argument could be used to justify all manner of injury to suspects since the scale of imaginable damage is limitless.

The potential scope of the impact on human subjects that is associated with data mining is also particularly troubling. Apart from the dangers of the unauthorized disclosure of personal data or use of the data for illicit purposes, many of the data mining programmes examined would involve broad collection and processing of data that in most instances would inevitably pertain to innocent individuals. This aspect of data mining will be explored in more detail with regard to the implications under international law in Deliverable D08.3. Yet, the implication is that broad application of data mining more or less places everyone engaged in certain legitimate behaviour within the scope of investigation without any suspicion of criminal activity. Some may respond to this argument by claiming that there is a kind of anonymity in the data flood. In other words, those conducting data mining exercises will only be interested in that data which shows up in the results, and most individuals whose data is subject to data mining will not receive extra scrutiny or otherwise be inconvenienced. We will discuss the legal merits of this argument in Deliverable 8.3. In any event, from an ethical standpoint, it still represents the handling and processing of personal data without any level of suspicion and exposes those individuals to a variety of further privacy violations as well as adverse administrative decisions as discussed above. More targeted applications of data mining that takes place within the context of an investigation, where the processing is carried out on data that was duly collected under appropriate legal standards, would avoid this aspect of indiscriminate data processing.

Despite the poor performance of the *Rasterfahndung*, we discovered some anecdotal evidence of applications of data mining that showed promise; and absent the availability of studies of data mining performance, we are also left to speculate on an intuitive basis as to what sort of approaches might prove more effective and limit negative impact in terms of unwarranted privacy infringement. For this purpose, we return to the typology of programmes that was introduced in Section 3.2 above.

Discovery of terrorists and terrorist networks – As noted above, one method which has been proposed for catching or uncovering terrorists is through the use of profiling. It is, however, unclear whether it is possible to come up with an effective profile or profiles of terrorists. Much attention has been devoted to al Qaeda and Islamic terrorism. This focus may perhaps be appropriate due to the relative level of threat, but it could also bring with it the enhanced risk of discrimination on the basis of religion, ethnicity or national origin. Such a system would have the result that individuals of South Asian or Arabic origin or of Muslim faith would be more likely to be subjected to increased scrutiny than others. Since false positives are inevitable, this fact would also mean that more innocent individuals of South Asian or Arabic origin or of Muslim faith would be targeted than would be the case for other

[204] The JASON Report on "Rare Events" estimates that the likelihood that another event on the scale of 9/11 or greater will happen before 2020 is about 7%. JASON/ MITRE Corporation, "Rare Events", JSR-09-108 (October 2009) at 28.

classes of innocents. Moreover, a profile that includes ethnicity or national origin as significant criteria can easily be circumvented by relying on operatives of a different ethnicity or national origin.[205] In perhaps the most obvious setting where computerized profiling has been applied—aviation screening—data indicating ethnicity or national origin will often not be available. Instead these systems would have to rely on factors such as the country of issue of the passenger's passport and travel patterns as proxies for these characteristics.

Behavioural profiling—which would include reliance on travel patterns—may pose less risk of discrimination on the basis of personal attributes (though there might still be a risk of indirect discrimination), but can be just as problematic in terms of infringement of the right to privacy. Here also, there must be a relatively high correlation between the selected behaviour and terrorist activity or affiliation with terrorists. Additionally, virtually every travel pattern could be linked to innocent behaviour. Therefore, the true challenge for the efficacy of travel pattern profiling becomes an issue of whether the innocent patterns can be discerned from the non-innocent ones.

Reliance on link analysis would appear useful in an investigatory setting when starting from a known suspect. The dangers with this approach are that it can lead to an assumption of guilt by association or be used to cast a wider net of intrusion than might be justified by the circumstances. Although particular emphasis has been placed on developing methods to discover previously unknown terrorist suspects or "sleepers", it is important to remember that one lesson from 9/11 was that many of the individuals involved in the plot were already known to intelligence agencies. With the exception of individuals who truly act alone of their own motives, even so-called "lone wolf" terrorists will generally have some contact with terrorist networks and often require it in order to acquire the capability to perform terrorist acts.

Profile generation – The use of data mining for the purposes of profile generation may be more effective than some of the other methods of profile generation. The *Rasterfahndung* profile and MATRIX terrorist factor were based on the assumption that other terrorists would share the same characteristics of a small set of terrorists associated with a single terrorist act. Additionally, the method of selecting which factors were relevant for the profile appeared to rely on an intuitive approach, which must be regarded as arbitrary and is fraught with the possibility that human bias may play a role in the selection. Data mining, on the other hand, when done properly, may hold the potential to discover commonalities among known terrorists in an empirical and unbiased manner. Nonetheless, it suffers from the same issues of whether the applied methodology is sound as is the case with the use of data mining to perform the profiling itself.

Risk assessment – Currently documented experience with the application of crime mapping and geospatial predictive analytics suggests that these approaches can prove effective for making decisions with respect to the allocation of resources. Additionally, these methods avoid privacy issues due to the absence of personal data: these approaches target areas rather than individuals.

[205] A 2005 undisclosed FBI report allegedly stated that already at that time "'Al-Qa'ida places a premium on operatives who are not, or at least appear not to be, Arab, particularly those with European or Asian features, according to various detainee reporting'". "Secret FBI Report Questions Al Qaeda Capabilities", ABC News, 09 March 2005., http://abcnews.go.com/WNT/Investigation/story?id=566425&page=2.

Communications analysis that aims to uncover communications relating to terrorist activity or planning specifically strikes us as a dubious approach. Although software designed to monitor employee communications in order to detect disclosure of trade secrets, intellectual property, or other forms of disloyalty likely provides a commercially available analogy, we have not been able to find any studies that assess the performance of such software. It is easy to conceive that programmes that rely on detecting the use of certain suspicious keywords would likely capture a large amount of innocent communication. Programmes which search for certain constellations of such keywords within a single message or a chain of message exchanges would likely prove more effective. However, here again the problem of finding a reliable profile or model of a terrorist message arises. Additionally, planners of terrorist acts may use coded language in order to conceal the actual content of their communications.

Preparatory acts monitoring represents a more targeted approach than, for example, air traveller profiling. Purchasing a certain combination of chemicals which could be used to make a bomb would seem to be more highly correlated to terrorist activity than a certain set of travel patterns. It would also be possible to establish the system in such a way that would minimize the impact on the right to privacy—namely by focusing on activities without consideration of the individuals involved. Once activities that could reasonably represent preparations for a terrorist act occurred, investigators could then seek to uncover the persons involved in order to identify suspects. From a practical standpoint, however, it is unclear how feasible such an approach would be since it would have to rely on a system of reporting where the reports would be submitted by, for instance, certain merchants. A model would be the system of SAR submissions on the part of US financial institutions.[206] However, any model that relied on combining more than one sort of activity—for instance purchases plus money transfers—would likely need to rely on identifying information in order to match the activities to the same person. An alternative to merchant reporting would be a system of mandatory registration, where the purchaser would have to tender some form of reliable identification and fill out a form that would be submitted to the appropriate authorities. Yet, such a system may be too burdensome for purchases involving common household products that contain potential bomb-making substances.

One area that has recently received significant attention from technology developers, security professionals, and critics is the use of automated video analytic technologies. Often the aim of such proposed programmes is to detect "suspicious" or "abnormal" behaviour or the emergence of such behaviour. Such systems may rely on machine learning to develop and improve profiles of suspicious behaviour and may depend on human input to "train" the system, where a user provides the system with feedback as to whether an event highlighted by the programme is indeed the kind of activity that the user is interested in. These exercises amount to a kind of profiling and face the same sort of problems mentioned above in relation to profiling technologies.[207] Machine learning, however, can present some

[206] For more information on the SAR system and regulatory background, see U.S. Government Accountability Office, "Suspicious Activity Report Use Is Increasing, but FinCEN Needs to Further Develop and Document Its Form Revision Process", GAO-09-226 (February 2009) at 11 et seq., http://www.gao.gov/new.items/d09226.pdf.

[207] See also A. Madrigal, "TSA Says Better Body Scanners Will End Privacy Uproar: Don't Bet on It", The Atlantic, 18 November 2010, http://www.theatlantic.com/technology/archive/2010/11/tsa-says-better-body-scanners-will-end-privacy-uproar-dont-bet-on-it/66761/, which describes the analogous difficulty of training a system to detect "suspicious" objects in the images produced by full-body scanners ("In order to make accurate determinations, they need a huge library of suspicious and normal images, said the Pacific Northwest National Laboratory's Doug McMakin, who developed the technology on which the L-3 SafeView system is based."). The body scanner algorithm deals with a

unique problems. Often the user and in some instance perhaps even the designers will not know or be able to determine what sort of rules the programme is developing to distinguish suspicious behaviour. Moreover, there is also an issue with regard to the level of sophistication of the learning mechanism. When the system highlights an event as suspicious or threatening and is then given input from the human operator that the event is not of interest, simply revising the last rule that the system established may not resolve the issue. It may be difficult to determine how much training the system will require before it learns to correct a rule that will generally produce correct results despite the fact that it is essentially off base. Deploying such systems in multicultural environments such as international airports is problematic since behaviour that is "abnormal" or unusual in one culture may be "normal" or common in another. Here also bias can creep in where the system is not trained on such diverse populations or the system simply inherits the bias of the trainer

Analytic aids – It is difficult to assess those programmes categorized as analytic aids due to the lack of information on them as well as our inability to identify commercially available analogues. Obvious dangers are that the aids will end up misleading analysts or that reliance on the aids will result in analysts failing to exercise critical thought.

6. Conclusion & Recommendations

Whilst there is little evidence demonstrating the effectiveness of counter-terrorism data mining, sweeping conclusions that disregard the particular context of different data mining programmes should be avoided. Whether a particular data mining programme is effective depends on a combination of numerous factors: what aims are sought, what methods are applied, the quality of the underlying data on which the analysis is performed, the amount and complexity of that data, the available processing power, the way in which the results are to be used, and the relative tolerance for false results. We have attempted to assess different applications of counter-terrorism data mining in different contexts, largely on an intuitive basis. However, comprehensive assessment of the methodology and approach behind a particular proposed application is critical. Where a particular application takes the form of profiling, significant care must be given to ensuring that the model on which the profile is based represents a reasonable approach and rests on sound methodology. In addition to testing of the efficacy of the profile, testing and assessment of the methods for updating the profile is also important. Beyond the issue of evaluation, we believe some approaches to data mining will be better at limiting the impact on privacy than others. For this reason we formulate a number of general recommendations concerning assessment of the technology and methods on the one hand and the limitation of the impact on the right to privacy on the other.[208]

4. Governments should require rigorous testing of data mining products before implementing them for counter-terrorism purposes.[209]

---

fixed image as opposed to the "moving" images of a video surveillance camera.

[208] The National Research Council has also developed a set of "Framework Criteria" for evaluating the effectiveness of data mining programmes. See National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment* (Washington, D.C: National Academies Press, 2008), pp. 59-60.

[209] DETECTER Deliverable D06.3 similarly recommends that the necessity of deploying a particular detection technology in the counter-terrorism context be "established and supported by concrete evidence". D06.3, Sec. 9.

4.1. Testing should be carried out on simulated data that reflects to the closest degree possible the forms and types of data to which the deployed product is intended to be applied.

In conducting initial testing to determine a programme's performance and efficacy, it is not necessary to utilize data relating to real persons. By refraining from the use of real personal data, agencies can avoid unwarranted violation of personal privacy or applicable law pertaining to the processing of personal data and avert the risks that the data is lost, stolen, or misused. Nonetheless, the data must reflect the types and character of data on which the programme is to be used in order for the testing to provide an accurate picture of the programme's rate of effectiveness in the target environment.

4.2. Testing should be performed by a competent body that is completely independent of the agency or agencies which would use the technology if it were to be deployed, and that body should possess the appropriate expertise in scientific evaluation and assessment methodology.

There a number of inherent dangers where an agency conducts its own internal testing. On the one extreme, assessment may merely take the form of a rubber-stamp process in which effectively no real assessment takes place. Toward the other end of the scale, there are the risks that conflicts of interest or internal pressures influence the design of the testing or the observations or conclusions that are drawn from it. Additionally, internal review of technology bears the risk that standards or benchmarks are established that would be deemed inappropriate or unsatisfactory by the broader community. Those individuals involved in testing data mining programmes should possess the requisite knowledge of the issues in and accepted methods for the evaluation of computational systems.

4.3. Ideally, results of tests involving deployed technologies should be published for public review so that citizens may assess whether the technology represents a worthwhile investment of public funds.

The argument is often made that disclosing the results of tests and assessments would pose a threat to national security by providing terrorists with information that could be used to circumvent the counter-terrorism measures in question. However, it is unclear how the publication of performance figures could undermine security measures. At best, such numbers could give terrorists some notion of the chance of success at slipping through the system, but they would not provide any insight as to how to evade the system.

4.4. Only those data mining programmes that can demonstrate their effectiveness in the test setting should be allowed to be deployed.

5. Following implementation, programmes should be reviewed on a regular basis and monitoring mechanisms updated accordingly.

Regular reviews can ensure that the system continues to function properly and effectively. Where automated monitoring against profiles takes place, those profiles should be updated regularly to reflect the current state of affairs and the emergence of new modi operandi.

6. Governments should establish parameters for the types of data that will be subject to data mining programmes or exercises.

During the design stage and after a sound methodology has been selected, parties responsible for designing a data mining system should determine what types of data are actually required for the programme to effectively perform its function.

6.1. Parameters should be based on the minimum amount and types of personal data necessary to conduct the analysis for the aims that it seeks to achieve.

In order to limit potential infringement of the right to privacy, the use of personal data should be avoided wherever possible. Moreover, governments should seek to minimize not only the amount of data that is to be subjected to data mining but also the number of different forms of data.[210]

6.2. Where a data mining programme is applied to personal data, use of the programme should be confined to the greatest extent possible to investigatory applications centring on known suspects and endeavour to comply with traditional standards governing government intrusion into the private life of individuals.

Limiting the application of data mining to known suspects could minimize the potential impact on innocent individuals. Investigatory applications of data mining in this context could include such things as the analysis of telephone call data, cash flows, and travel patterns.

[210] The potential to utilize technological measures to preserve privacy within data mining programmes as well as particular categories of data that are owed particular deference will be discussed in D08.3.